The quantile spectral density and comparison based tests for nonlinear time series

Junbum Lee and Suhasini Subba Rao
Department of Statistics, Texas A&M University
College Station, U.S.A.
jlee@stat.tamu.edu and suhasini@stat.tamu.edu

Abstract

In this paper we consider tests for nonlinear time series, which are motivated by the notion of serial dependence. The proposed tests are based on comparisons with the quantile spectral density, which can be considered as a quantile version of the usual spectral density function. The quantile spectral density 'measures' the sequential dependence structure of a time series, and is well defined under relatively weak mixing conditions. We propose an estimator for the quantile spectral density and derive its asympototic sampling properties. We use the quantile spectral density to construct a goodness of fit test for time series and explain how this test can also be used for comparing the sequential dependence structure of two time series. The asymptotic sampling properties of the test statistic is derived under the null and an alternative. Furthermore, a bootstrap procedure it proposed to obtain a finite sample approximation. The method is illustrated with simulations and some real data examples.

Key words and phrases: Bootstrap, goodness of fit tests, mixing, nonlinear time series, quantile spectral density.

1 Introduction

The analysis of most time series is based on a set of assumptions, which in practice need to be tested. This is usually done through a goodness of fit test. The majority of goodness of fit tests for time series are based on fitting the conjectured model to the data, estimating the residuals of the model and testing for lack of correlation, normally with a Ljung-Box type test (see for example, Anderson (1993) and Hong (1996)). Chen and Deo (2004) propose a test based on the spectral density, and Hallin and Puri (1992) propose robust tests based on ranks. If one restricts

the class of models to just linear time series models, then these tests can correctly identify the model. However, problems can arise, if one widens the class of models and allow for nonlinear time series. For example, if the time series were to satisfy an ARCH process, then it will be uncorrelated, but it is not independent. Furthermore, the squares will satisfy an autoregressive representation, with errors which are martingale differences. Therefore, correlation based test for nonlinear time series models may not identify the model.

Neumann and Paparoditis (2008) propose a goodness of fit test for Markov time series models based on the one step ahead transition distribution. But this test is specifically for Markov models. An alternative approach is to generalise the notion of correlation to measuring the general dependence between pairs of random variables in a time series. This notion is usually called serial dependence, and can be traced dates back to Hoeffding (1948). Skaug and Tjøstheim (1993) and Hong (2000) use this definition to test for serial independence of a time series. Hong (1998) takes these notions further, and generalises the spectral density to sequential dependence. He does this by defining the generalised spectral density, which is the Fourier transform of the characteristic function of pair-wise dependent data. He uses this device in Hong (1998) and Hong and Lee (2003) to test for goodness of fit of a time series model, mainly through the analysis of the estimated residuals. However, sometimes the residuals cannot be or are not easy to estimate. For example, it is possible to estimate the residuals of an ARCH ($X_t = Z_t\sigma_t$), possible but difficult with a GARCH and usually impossible for many models of the type $X_t = g(X_{t-1}, \varepsilon_t)$.

In this paper, we use the notion of serial dependence to test for goodness of fit, but without estimating the residuals. In Section 2.1 we motivate our test by considering the Microsoft daily log return data and compare it with the GARCH$(1, 1)$ model (one of the standard models for such data sets). We show that though the GARCH model seems to model well some of the stylised facts of this data, ie. the uncorrelatedness of the data and positive correlation in the absolute values, if one made a deeper analysis and compared the correlation of other transformations such as $\mathrm{cov}(I(X_t \leq x), I(X_{t+r} \leq y))$ (where I denotes the indictator function), there is large difference between the data and GARCH model. This motivates us to define the *quantile* autocovariance function and the *quantile* spectral density. The quantile spectral density can be considered as a measure of serial dependence of a time series. In Sections 2.2 and 2.3 we propose a method for estimating the quantile spectral density, and use the quantile spectral density as the basis of a test which compares the quantile spectral density estimator with the spectral density estimator under the null hypothesis. The asymptotic sampling properties of the quantile spectral density estimator are derived in Section 3.1. Recently there have been several articles defining and estimating the spectral density of sequential dependence. In particular, Li (2008), Hagemann (2011) and Dette, Hallin, Kley, and Volgushav (2011) define spectral density functions similar to the quantile spectral density, however these authors, estimate the periodogram and the quantile spectral density using L_1 methods. In contrast, our approach is motivated by the definition of the periodogram, this leads to an estimator of the quantile spectral density with an analytic

form, thus can easily be used in both goodness of fit and other tests. However, it is interesting, and rather surprising, to note that the L_1 estimator proposed in Dette et al. (2011) and our estimator of the quantile spectral density share similar asymptotic properties. In Section 3.2 we derive the asymptotic sampling properties of the test statistic. An advantage of our approach is that it can easily be extended to test other quantities, for example with a small adaption it can be used to test for equality of serial dependence of two time series, this is considered in Section 4. In Section 5 we propose a bootstrap method for estimating the finite sampling distribution of the test statistic under the null. The proofs can be found in the Appendix and technical report.

2 The quantile spectral density and the test statistic

2.1 Motivation

To motivate our approach, we analyze the Microsoft daily log returns (MSFT) between March 1986 - June 2003, which we denote as $\{X_t\}$. One argument for fitting GARCH types models to financial data is their ability to model the so called 'stylised facts' seen in such data sets. We now demonstrate why this is the case for the MSFT data (see Zivot (2009)). Using the maximim likelihood, the GARCH model which best fits the log differences of the MSFT is $X_t = \mu + \varepsilon_t$, where $\epsilon_t = \sigma_t Z_t$, $\sigma_t^2 = a_0 + a_1\varepsilon_{t-1}^2 + b\sigma_{t-1}^2$ ($\{Z_t\}$ are independent, identically distributed standard normal random variables), with $\mu = 1.56 \times 10^{-3}$, $a_0 = 1.03 \times 10^{-5}$, $a_1 = 0.06$ and $b = 0.925$. In Figure 1 we give the sample autocorrelation plots of $\{X_t\}$ and $\{|X_t|\}$, together with the autocorrelation plots of the corresponding GARCH$(1,1)$ model. Comparing the two plots, it appears that the GARCH$(1,1)$ captures the 'stylised facts' in the Microsoft data, such as the near zero autocorrelation of the observations and the persistant positive autocorrelations of the absolute log returns. However, if we want to check the suitability of the GARCH model for modelling the general pair-wise dependence structure, that is the joint distribution of (X_s, X_t) for all s and t (often called *sequential dependence*), then we need to look beyond the covariance of $\{X_t\}$ and $\{|X_t|\}$. To make a more general comparison we transform the data into indicator variables $\{I(X_t \leq x)\}$ and check the correlation structure of the indicator variables over various x. For example, define the multivariate vector time series $\underline{Y}_t = (I(X_t \leq q_{0.1}), I(X_t < q_{0.5}), I(X_t \leq q_{0.9}))$, where q_α denotes the estimated α-percentile of X_t. Plots of the cross-covariances of $\underline{Y}_t$ and the corresponding GARCH model (with Gaussian innovations) are given in Figure 2. In Figure 2, there are clear differences in the dependence structure of the data and the GARCH model. The 10th, 50th and 90th percentiles correspond to large negative, zero and large positive values of X_t (big negative change, no change and large positive changes in the returns). In order to do the analysis, we will use the following observations. By using that $\mathrm{cov}(I(X_0 \leq x), I(X_r \leq y)) =$

$P(X_0 \leq x, X_r \leq y) - P(X_0 \leq x)P(X_r \leq y)$, for all $x, y \in \mathbb{R}$ we have

$$\mathrm{cov}(I(X_0 \leq x), I(X_r \leq y)) = \mathrm{cov}(I(X_0 \geq x), I(X_r \geq y)) = -\mathrm{cov}(I(X_0 \leq x), I(X_r > y)).$$

From Figure 2 we observe:

- The ACF of $I(X_t \leq q_{0.5})$ of the GARCH is zero. This is due to the symmetry of the GARCH process: given the event $X_0 \leq 0$, we have equal chance $X_r > 0$ and $X_r < 0$ (ie. $\mathrm{cov}(I(X_0 \leq 0), I(X_r \leq 0)) = -\mathrm{cov}(I(X_0 \leq 0), I(X_r > 0)))$. This means that $\mathrm{cov}(I(X_0 \leq 0), I(X_r \leq 0)) = 0$. On the other hand, for the MSFT data we see that there is a clear positive correlation in the sample autocorrelation of $\{I(X_t < 0)\}$. One interpretation for this behaviour, is that a decrease in consecutive values, is likely to lead to future decreases.

- The cross correlation $\mathrm{cov}(I(X_0 < q_{0.1}), I(X_r < q_{0.9}))$, where $\{X_t\}$ comes from a GARCH is symmetric, ie. $\mathrm{cov}(I(X_0 < q_{0.1}), I(X_r < q_{0.9})) = \mathrm{cov}(I(X_0 < q_{0.1}), I(X_{-r} < q_{0.9}))$. On the other hand, the corresponding sample cross-correlations of the MSFT is not symmetric. Thus the GARCH process is time reversible, whereas it appears that the MSFT data may not be.

The cross and autocovariances in Figure 2 are a graphical representation of the serial dependence structure of the time series. These plots suggest that for the MSFT data the GARCH model may not be the most appropriate model, especially if validity is based on modelling the serial dependence structure. In the sections below we will test this.

2.2 The quantile spectral density function

We now formalise the discussion above. Let us suppose that $\{X_t\}$ is a strictly stationary time series. It is obvious that the cross covariance of the indicator functions $\{I(X_t \leq x), I(X_t \leq y)\}$ is

$$C_r(x,y) \ := \ \mathrm{cov}(I(X_0 \leq x), I(X_r \leq y)) = P(X_0 \leq x, X_r \leq y) - P(X_0 \leq x)P(X_r \leq y).$$

Skaug and Tjøstheim (1993) and Hong (2000) use a similar quantity to test for serial independence of a time series. We call $C_r(\cdot)$ the *quantile covariance*. If $\{X_t\}$ is an α-mixing time series with mixing rate $s > 1$ (s is defined in Assumption 3.1, below) it can be shown that $\sup_{x,y} \sum_r |\mathrm{cov}(I(X_0 \leq x), I(X_r \leq y))| < \infty$, thus for all $x, y \in \mathbb{R}$, it's Fourier transform

$$G(x,y;\omega) \ = \ \frac{1}{2\pi} \sum_r C_r(x,y;\omega) \exp(ir\omega),$$

is well defined. Since $G(x, y; \omega)$ can be considered as the cross-spectral density of $\{I(X_t < x), I(X_t < y)\}$, we call $G(\cdot)$ the *quantile spectral density.*

2.2.1 Properties of the quantile spectral density

The quantile spectral density carries all the information about the serial dependence structure of the time series. For example (i) if $\{X_t\}$ is serially independent, then G does not depend on ω and $G(x, y; \omega) = \rho(x, y)$ (ii) if for all r, the distribution function of (X_0, X_r) is identical to the distribution function of (X_0, X_{-r}), then $G(\cdot)$ will be real and (iii) for any given x and y, G gives information about any periodicities that may exists at a given threshold. In addition, $G(\cdot)$ captures the covariance structure of any transformation of $\{X_t\}$. For example, consider the transformation $\{h(X_t)\}$, then it is straightforward to show that the spectral density of the time series $\{h(X_t)\}$ is

$$f_h(\omega) = \frac{1}{2\pi} \sum_r \mathrm{cov}(h(X_0), h(X_r)) \exp(ir\omega) = \int \int h(x)h(y)G(dx, dy; \omega).$$

Of course, $G(x, y; \omega)$ only captures the serial dependency, and may miss higher order structure. Only in the case that $\{X_t\}$ is Markovian, does $G(x, y; \omega)$ capture the entire joint distribution of $\{X_t\}$.

Remark 2.1 *The quantile spectral density is closely related to the generalised spectral density introduced in Hong (1998). He defines the generalised spectral density as* $h(x, y; \omega) = \sum_r cov(\exp(ixX_0), \exp(iyX_r)) \exp(ir\omega)$. *Essentially, this is the Fourier transform of the characteristic function of pairwise distributions minus their marginals, therefore the relationship between the quantile spectral density and the generalised spectral density is analogous to that between the distribution function and the characteristic function of a random variable. Hong (1998, 2003) uses the generalised spectral density as a tool in various tests goodness of fit tests, which are mainly based on the residual. On the other hand, the goodness of fit test that we propose, is based on checking for similarity between the estimated quantile spectral density and the proposed spectral density.*

Remark 2.2 (The Copula spectral density) *A closely related quantity to the quantile spectral density is the copula spectral density, which is defined as*

$$G_{\mathcal{C}}(u_1, u_2; \omega) \quad = \quad \frac{1}{2\pi} \sum_r \mathcal{C}_r(u_1, u_2; \omega) \exp(ir\omega), \tag{1}$$

where $\mathcal{C}_r(u_1, u_2) = cov(I(F(X_0) \leq u_1), I(F(X_r) \leq u_2)) = \mathbb{E}(I(F(X_0) \leq u_1)I(F(X_r) \leq u_2)) - u_1u_2$, *and* $F(\cdot)$ *is marginal distribution function of* $\{X_t\}$. *Note that by definition* $u_1, u_2 \in [0, 1]$.

Thus, unlike the quantile spectral density, the copula spectral density is invariant to any monotonic transformation of $\{X_t\}$, for example mean and variance shifts. By considering the ranks of $\{X_t\}$, the methods detailed in the section below can also be used to estimate $G_{\mathcal{C}}$. Dette et al. (2011) have recently proposed L_1-methods for estimating $G_{\mathcal{C}}$, and the asymptotic sampling properties have been derived for this estimator.

In Figures 3, 4 and 5 we plot the quantile spectral density for the autoregressive ($X_t = 0.9X_{t-1} + Z_t$), ARCH ($X_t = \sigma_t Z_t$ with $\sigma_t^2 = 1/1.9 + 0.9X_{t-1}^2$) and squared ARCH, with independent, identically distributed (iid) Gaussian innovations Z_t. The diagonals are of $G(x, x; \omega)$, the lower triangle contains the real part of $G(x, y; \omega)$ and the upper triangle the imaginary part of $G(x, y; \omega)$. We observe that the AR and ARCH quantile spectral densities are very different. The AR has a similar shape for all x, whereas for the ARCH, it is flat (like the spectral density of uncorrelated data) at about the 50% percentile, but moves away from flatness at the extremes. Furthermore, recalling that the AR and ARCH squared have the same spectral density (if the moments of the ARCH squared exists), there is a large difference between the quantile spectral density of the AR and the ARCH squared.

2.2.2 Estimating the quantile spectral density

The quantile spectral density $G(x, y; \omega)$ can be considered as the cross spectral density of the bivariate time series $\{I(X_t \leq x), I(Y_t \leq y)\}$. Therefore, our estimator of $G(x, y; \omega)$ is motivated by the classical cross spectral. To do this we define the class of lag windows we shall use.

Definition 2.1 *The lag window takes the form*

$$\lambda(u) = \Big(\sum_{j=-r}^{r} a_r \exp(i2\pi ru) - \sum_{j=1}^{r} b_j |u|^j\Big) I_{[-1,1]}(u),$$

where $I_{[-1,1]}(u) = 1$ for $u \in [-1, 1]$ and zero otherwise. This class of lag windows is quite large, and includes the truncated window, the Bartlett window and general Tukey window (see, for example, Priestley (1981) Section 6.2.3 for properties of these lag windows).

To obtain an estimator of G, we define the centralised, transformed variable $Z_t(x) = I(X_t \leq x) - \widehat{F}_T(x)$ (where $\widehat{F}_T(x) = \frac{1}{T}\sum_t I(X_t \leq x)$). We estimate the quantile covariance $C_r(x, y) = P(X_0 \leq x, X_r \leq y) - P(X_0 \leq x)P(X_r \leq y)$ with $\widehat{C}_r(x, y) = \frac{1}{T}\sum Z_t(x)Z_{t+r}(y)$, and use as an estimator of G

$$\begin{aligned} \widehat{G}_T(x, y; \omega_k) &= \frac{1}{2\pi}\sum_r \lambda_M(r)\widehat{C}_r(x, y)\exp(ir\omega_k) \\ &= \sum_s K_M(\omega_k - \omega_s) J_T(x; \omega_s)\overline{J_T(y; \omega_s)}, \end{aligned} \tag{2}$$

where $\lambda_M(r) = \lambda(r/M)$, $J_T(x;\omega) = \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T} Z_t(x)\exp(it\omega)$ and $K_M(\omega) = \frac{1}{T}\sum_r \lambda_M(r)\exp(ir\omega)$.

2.3 The test statistic

The proposed test is based on the fit of the estimated quantile spectral density to the conjectured quantile spectral density. More precisely, we test $H_0 : G(x,y;\omega) = G_0(x,y;\omega)$ against $H_A : G(x,y;\omega) \neq G_0(x,y;\omega)$, where G is the quantile spectral density of $\{X_t\}$, $G_0(x,y;\omega) = \frac{1}{2\pi}\sum_r C_{0,r}(x,y)\exp(ir\omega)$ and $C_{0,r}(x,y) = F_{0,r}(x,y) - F_0(x)F_0(y)$. Thus under the null the marginal distribution is $F_0(\cdot)$ and the joint distribution is $F_{0,r}(\cdot)$. We use the quadratic distance to measure the distance between the estimated quantile spectral density and the conjectured spectral density, and define the test statistic as

$$\begin{aligned} \mathcal{Q}_T &= \frac{1}{T}\sum_{k=1}^{T}\int |\widehat{G}_T(x,y;\omega_k) - \frac{1}{2\pi}\sum_r \lambda_M(r)C_{0,r}(x,y)\exp(ir\omega_k)|^2 dF_0(x)dF_0(y) \\ &= \frac{1}{T}\sum_{k=1}^{T}\int |\widehat{G}_T(x,y;\omega_k) - \sum_{s=1}^{T} K_M(\omega_k - \omega_s)G_0(\omega_s)|^2 dF_0(x)dF_0(y) \\ &= \frac{1}{2\pi}\sum_r \lambda_M(r)^2 \int\int \left|\hat{C}_r(x,y) - C_{0,r}(x,y)\right|^2 dF_0(x)dF_0(y), \end{aligned} \tag{3}$$

where the above immediately follows from Parseval's theorem. The choice of lag window will have an influence on the type of alternatives the test can detect. For example, the truncated window ($\lambda(u) = I_{[-1,1]}(u)$) gives equal weights to all the quantile covariances, whereas the Bartlett window ($\lambda(u) = (1-|u|)I_{[-1,1]}(u)$) gives more weight to the lower order lags. Therefore the tests ability to detect the alternative will depend on which lags of the quantile covariance deviates the most from the null, and the weight the lag window places on these. We derive the asymptotic distribution of $\mathcal{Q}_T$ in Section 3.2.

Remark 2.3 *The test can be adapted to be invariant to monotonic transformations (such as shifts of mean and variance). This can be done by replacing the quantile spectral density with the copula spectral density $G_{\mathcal{C}}(\cdot)$ defined in (1). In this case the null is $H_0 : G_{\mathcal{C}}(x,y;\omega) = G_{\mathcal{C},0}(x,y;\omega) = \frac{1}{2\pi}\sum_r \mathcal{C}_{0,r}(u_1,u_2;\omega)\exp(ir\omega)$ against $H_A : G_{\mathcal{C}}(x,y;\omega) \neq G_{\mathcal{C},0}(x,y;\omega)$. The test statistic in this case is*

$$\mathcal{Q}_{T,\mathcal{C}} = \frac{1}{T}\sum_{k=1}^{T}\int |\widehat{G}_{T,\mathcal{C}}(u_1,u_2;\omega_k) - \frac{1}{2\pi}\sum_r \lambda_M(r)\mathcal{C}_{0,r}(u_1,u_2)\exp(ir\omega_k)|^2 du_1 du_2,$$

where we estimate $\widehat{G}_{T,\mathcal{C}}(u_1,u_2;\omega_k)$ in the same way as we have estimated $\widehat{G}_T$ in (2) but replace $\{X_t\}_t$ with $\{\hat{F}_T(X_t)\}_t$. The distribution of $\mathcal{Q}_{T,\mathcal{C}}$ is beyond the scope of the current paper.

3 Sampling properties

In this section we derive the sampling properties of the quantile spectral density $\widehat{G}_T$ and the test statistic $\mathcal{Q}_T$. We will use the α-mixing assumptions below.

Assumption 3.1 *Let us suppose that $\{X_t\}$ is a strictly stationary α-mixing time series such that*

$$\sup_{\substack{A\in\sigma(X_r,X_{r+1},\ldots)\\ B\in\sigma(X_0,X_{-1},\ldots)}} |P(A\cap B)-P(A)P(B)| \leq \alpha(r),$$

where $\alpha(r)$ are the mixing coefficients which satisfy $\alpha(r)\leq K|r|^{-s}$ for some $s>2$.

3.1 Sampling properties of $\widehat{G}_T$

In the following lemma we derive the limiting distribution of $\widehat{G}_T$, this will allow us to construct point wise confidence intervals for G.

Theorem 3.1 *Suppose Assumption 3.1 holds. Then*

$$\mathbb{E}(\widehat{G}_T(x,y;\omega)) = G(x,y;\omega) + O(\frac{1}{M^{s-1}}),$$

and for $0<\omega_k<\pi$ we have

$$\begin{aligned} V_T(x,y;\omega_k)^{-1/2}\left(\begin{array}{c} \Re\widehat{G}_T(x,y;\omega_k) - \Re\mathbb{E}(\widehat{G}_T(x,y;\omega_k)) \\ \Im\widehat{G}_T(x,y;\omega_k) - \Im\mathbb{E}(\widehat{G}_T(x,y;\omega_k)) \end{array}\right) &\stackrel{D}{\to} \mathcal{N}\big(0,I_2\big) \\ V_T(x,x;\omega_k)^{-1/2}\Big(\widehat{G}_T(x,x;\omega_k) - \mathbb{E}(\widehat{G}_T(x,x;\omega_k))\Big) &\stackrel{D}{\to} \mathcal{N}(0,1), \end{aligned}$$

where $M\to\infty$ and $M/T\to 0$ as $T\to\infty$,

$$V_T(x,y;\omega_k) = \sum_{k=1}^{T} K_M(\omega_k-\omega_s)^2 \left(\begin{array}{cc} A(x,y;\omega_s) & C(x,y;\omega_s) \\ C(x,y;\omega_s) & B(x,y;\omega_s) \end{array}\right) = O(\frac{M}{T}),$$

and

$$\begin{aligned} A(x,y;\omega_s) &= \frac{1}{2}\Big(G(x,x;\omega_s)G(y,y;\omega_s) + \Re G(x,y;\omega_s)^2 - \Im G(x,y;\omega_s)^2\Big) \\ B(x,y;\omega_s) &= \frac{1}{2}\Big(G(x,x;\omega_s)G(y,y;\omega_s) + \Im G(x,y;\omega_s)^2 - \Re G(x,y;\omega_s)^2\Big) \\ C(x,y;\omega_s) &= \Re G(x,y;\omega_s)\Im G(x,y;\omega_s). \end{aligned}$$

Thus, if $\frac{M}{T} >> \frac{1}{M^{2(s-1)}}$, in other words the variance of $\widehat{G}_T$ dominates the bias, then we can use the above result to construct confidence intervals for G.

3.2 Sampling properties of test statistic under the null hypothesis

We now derive the limiting distribution of the test statistic under the null hypothesis. Let

$$\begin{aligned} E_T &= \frac{1}{T}\int\int W_M(\omega-\theta)^2 G(x,x;\theta)G(y,y;\theta)dF_0(x)dF_0(y)d\theta d\omega \\ V_T &= \frac{4}{T^2}\int\int \Delta_M(\theta_1-\theta_2)^2\prod_{i=1}^{2}G(x_1,x_2;\theta_i)G(y_1,y_2;\theta_i)d\theta_i dF_0(x_i)dF_0(y_i), \end{aligned} \tag{4}$$

where

$$\begin{aligned} W_M(\theta) = \frac{T}{2\pi}K_M(\theta) &= \frac{1}{2\pi}\sum_r \lambda_M(r)\exp(ir\theta) \\ \Delta_M(\theta_1-\theta_2) &= \int W_M(\omega-\theta_1)W_M(\omega-\theta_2)d\omega. \end{aligned} \tag{5}$$

Lemma 3.1 *Suppose that Assumption 3.1 holds and $G(\cdot)$ is the quantile spectral density of $\{X_t\}$. Then under the null hypthesis we have*

$$\mathbb{E}\big(\mathcal{Q}_T\big) = E_T + O(\frac{1}{T}) = O(\frac{M}{T}) \text{ and } var\big(\mathcal{Q}_T\big) = V_T + O(\frac{1}{T}) = O(\frac{M}{T^2}).$$

Using the above we obtain the limiting distribution under the null.

Theorem 3.2 *Suppose that Assumption 3.1 holds. Then under the null hypthesis we have*

$$V_T^{-1/2}\big(\mathcal{Q}_T - E_T\big) \xrightarrow{D} \mathcal{N}(0,1)$$

as $M\to\infty$ and $M/T\to 0$ as $T\to\infty$.

Using estimates of $\widehat{G}_T(\cdot)$, E_T and V_T can both be estimated. Thus by using the above result, we reject the null at the $\alpha\%$ level if $V_T^{-1/2}\big(\mathcal{Q}_T - E_T\big) > z_{1-\alpha}$ (where $z_{1-\alpha}$ denotes the $1-\alpha$ quantile of a standard normal distribution).

3.3 Behaviour of the test statistic under the alternative hypothesis

We now examine the behaviour of the test statistic under the alternative $H_A : G(x,y;\omega) = G_1(x,y;\omega) = \frac{1}{2\pi}\sum_r \big(F_{r,1}(x,y) - F_1(x)F_1(y)\big)\exp(ir\omega)$. To obtain the limiting distribution we

decompose the test statistic $\mathcal{Q}_T$ as $\mathcal{Q}_T = \mathcal{Q}_{T,1} + \mathcal{Q}_{T,2} + \mathcal{Q}_{T,3}$, where

$$\begin{aligned}
\mathcal{Q}_{T,1} &= \frac{1}{T}\sum_{k=1}^{T}\int \left|\widehat{G}_T(x,y;\omega_k) - \mathbb{E}(\widehat{G}_T(x,y;\omega_k))\right|^2 dF_0(x)dF_0(y) \\
\mathcal{Q}_{T,2} &= \frac{2}{T}\Re\sum_{k=1}^{T}\int \left[\hat{G}_T(x,y;\omega_k) - \mathbb{E}(\hat{G}_T(x,y;\omega_k))\right]\left[\mathbb{E}(\hat{G}_T(x,y;\omega_k)) - \tilde{G}(x,y;\omega_k)\right] dF_0(x)dF_0(y) \\
\mathcal{Q}_{T,3} &= \frac{1}{T}\sum_{k=1}^{T}\int \left|\mathbb{E}(\hat{G}_T(x,y;\omega_k)) - \tilde{G}(x,y;\omega_k)\right|^2 dF_0(x)dF_0(y),
\end{aligned}$$

and

$$\tilde{G}(x,y;\omega_k) = \frac{1}{2\pi}\sum_r \lambda_M(r)C_{r,0}(x,y)\exp(ir\omega) = \sum_s K_M(\omega_k - \omega_s)G_0(x,y;\omega_s).$$

From the decomposition of $\mathcal{Q}_T$, we observe that there are two stochastic terms $\mathcal{Q}_{T,1}$ and $\mathcal{Q}_{T,2}$, and a deterministic term $\mathcal{Q}_{T,3}$. By using Lemma 3.1, it can be shown that $\mathcal{Q}_{T,1} = O_p(\frac{M^{1/2}}{T} + \frac{M}{T})$. On the other hand, we show in the proof of the theorem below that $\mathcal{Q}_{T,2}$, is of lower order than $\mathcal{Q}_{T,1}$ and, thus, determines the distribution of $\mathcal{Q}_T$. To understand the role that $\mathcal{Q}_{T,3}$ plays in the test, we replace $\tilde{G}(x,y;\omega)$ and $\mathbb{E}(\widehat{G}_T(x,y;\omega))$ with G_0 and G_1 respectively and obtain

$$\mathcal{Q}_{T,3} = \frac{1}{T}\sum_{k=1}^{T}\int \left|G_1(x,y;\omega_k) - G_0(x,y;\omega_k)\right|^2 dF_0(x)dF_0(y) + O(\frac{1}{M^{s-1}}).$$

Thus $\mathcal{Q}_{T,3}$ measures the deviation of the alternative from the null hypothesis, and shifts the mean of the test statistic.

Theorem 3.3 *Suppose that Assumption 3.1 holds, and for all* r, $\sup_{x,y}|C_{0,r}(x,y)| \le K|r|^{-(2+\delta)}$, *for some* $\delta > 0$. *Under the alternative hypothesis we have*

$$\sqrt{T}\mathcal{Q}_{T,2} \xrightarrow{D} \mathcal{N}(0, V_{T,2}), \tag{6}$$

and

$$\sqrt{T}\big(\mathcal{Q}_T - \mathcal{Q}_{T,3}\big) \xrightarrow{D} \mathcal{N}(0, V_{T,2}) \tag{7}$$

where $M \to \infty$ and $\sqrt{M}/T \to 0$ as $T \to \infty$, and

$$\begin{aligned} V_{T,2} &= \frac{8}{T}\Re \int\int \Lambda_T(x_1,y_1;\omega)\overline{\Lambda_T(x_2,y_2;\omega)} \\ &\quad \left\{G_1(x_1,x_2;\omega)G_1(y_1,y_2;\omega) + G_1(x_1,y_2;\omega)G_1(y_1,x_2;\omega)\right\}d\omega\prod_{i=1}^{2}dF_0(x_i)dF_0(y_i) \\ &\quad + \frac{8}{T}\Re \int\int \Lambda_T(x_1,y_1;\omega_1)\overline{\Lambda_T(x_2,y_2;\omega_2)}G_{(x_1,y_1,x_2,y_2)}(\omega_1,-\omega_1,\omega_2)\prod_{i=1}^{2}dF_0(x_i)dF_0(y_i)d\omega_i, \end{aligned}$$

where $\Lambda_T(x,y;\omega_s) = \frac{1}{2\pi}\sum_r \lambda_M(r)^2(\frac{T-|r|}{T})[C_{1,r}(x,y) - C_{0,r}(x,y)]\exp(ir\omega_k)$ and $G_{(x_1,y_1,x_2,y_2)}$ is the cross tri-spectral density of $\{(I(X_t \le x_1), I(X_t \le y_1), I(X_t \le x_2), I(X_t \le y_2))\}_t$.

The theorem above tells us that the mean of the test statistic is shifted the further the alternative is from the null. Interestingly, we observe from the definition of $\Lambda_T(\cdot)$, that the variance also depends on the difference between the null and alternative. However, for a fixed alternative, the power of the test converges to 100% as the sample size grow.

4 Testing for equality of serial dependence of two time series

The above test statistic can easily be adapted to test other hypothesis. In this section, we consider one such example, and test for equality of serial dependence between two time series. Let us suppose that $\{U_t\}$ and $\{V_t\}$ are two stationary time series, and we wish to test whether they have the same sequential dependence structure. Using the same motivation as that for the the goodness of fit test described above we define the test statistic

$$\mathcal{P}_T = \frac{1}{T}\sum_{k=1}^{T}\int |\widehat{G}_{1,T}(x,y;\omega_k) - \widehat{G}_{2,T}(x,y;\omega_k)|^2 dF(x)dF(y),$$

where $\widehat{G}_{1,T}$ and $\widehat{G}_{2,T}$ are the quantile spectral density estimators based on $\{U_t\}$ and $\{V_t\}$ respectively and F is any distribution function. In order to obtain the limiting distribution under the null hypothesis we have $H_0 : G_1(x,y;\omega) = G_2(x,y;\omega)$ and the alternative $H_A : G_1(x,y;\omega) \neq G_2(x,y;\omega)$ we expand $\mathcal{P}_T$

$$\mathcal{P}_T := \mathcal{Q}_{1,1,T} + \mathcal{Q}_{2,2,T} - \mathcal{Q}_{1,2,T} - \mathcal{Q}_{2,1,T} + 2\mathcal{L}_{1,T} + 2\mathcal{L}_{2,T} + \mathcal{D},$$

where

$$\mathcal{Q}_{i,j,T} = \frac{1}{T}\sum_{k=1}^{T}\int \big[\widehat{G}_{i,T}(x,y;\omega_k) - \mathbb{E}(\widehat{G}_{i,T}(x,y;\omega_k))\big]\big[\overline{\widehat{G}_{j,T}(x,y;\omega_k) - \mathbb{E}(\widehat{G}_{j,T}(x,y;\omega_k))}\big]dF(x)dF(y),$$

$$\mathcal{L}_{i,T} = \Re\frac{1}{T}\sum_{k=1}^{T}\int \big[\widehat{G}_{i,T}(x,y;\omega_k) - \mathbb{E}(\widehat{G}_{i,T}(x,y;\omega_k))\big]\big[\mathbb{E}(\widehat{G}_{1,T}(x,y;\omega)) - \mathbb{E}(\widehat{G}_{2,T}(x,y;\omega))\big]dF(x)dF(y)$$

and

$$\mathcal{D} = \int\int\int |\mathbb{E}(\widehat{G}_{1,T}(x,y;\omega)) - \mathbb{E}(\widehat{G}_{2,T}(x,y;\omega))|^2 dF(x)dF(y)d\omega.$$

Therefore, using the above expansion under the null hypothesis we have

$$\mathcal{P}_T := \mathcal{Q}_{1,1,T} + \mathcal{Q}_{2,2,T} - \mathcal{Q}_{1,2,T} - \mathcal{Q}_{2,1,T},$$

where the moments are $\mathbb{E}\big(\mathcal{P}_T\big) = E_{T,3} + O(\frac{1}{T}) = O(\frac{M}{T})$ and $\mathrm{var}\big(\mathcal{P}_T\big) = V_{T,3} + O(\frac{1}{T}) = O(\frac{M}{T^2})$, with

$$\begin{aligned} E_{T,3} &= \frac{1}{T}\int\int W_M(\omega-\theta)^2\big(G_1(x,x;\theta)G_1(y,y;\theta) + G_2(x,x;\theta)G_2(y,y;\theta)\big)dF(x)dF(y)d\theta d\omega \\ V_{T,3} &= \frac{4}{T^2}\sum_{i=1}^{2}\int\int \Delta_M(\theta_1-\theta_2)^2\prod_{j=1}^{2}G_i(x_1,y_2;\theta_i)G_j(y_1,x_2;\theta_j)d\theta_j dF(x_j)dF(y_j). \end{aligned}$$

By using identical arguments as those used in the proof of Theorem 3.2, under the null hypothesis we have

$$V_{T,3}^{-1/2}\big(\mathcal{P}_T - E_{T,3}\big) \xrightarrow{D} \mathcal{N}(0,1).$$

Using the above result, we test for equality of sequential dependence, that is we reject the null hypothesis at the α-level if $|V_{T,3}^{-1/2}(\mathcal{P}_T - E_{T,3})| > z_{1-\alpha}$.

The limiting distribution of the alternative can be derived using the same methods as those used to derive the limiting distribution of $\mathcal{Q}_T$ under its alternative. It can be shown that

$$\mathcal{P}_T - \mathcal{D} := \underbrace{2\mathcal{L}_{1,T} + 2\mathcal{L}_{2,T}}_{O_p(\frac{1}{\sqrt{T}})} + O_p(\frac{M^{1/2}}{T}),$$

where $2\mathcal{L}_{1,T} + 2\mathcal{L}_{2,T}$ can be approximated by a quadratic form. Using this quadratic approximation, asymptotic normality of the above can be shown. Thus under a fixed alternative the power

grows to 100% as $T \to \infty$.

Remark 4.1 *We can easily adapt our method to test that the distributions of* (X_0, X_r) *and* (X_{-r}, X_0) *are identical for all* r *(ie.* $F_r(x,y) = F_{-r}(x,y)$*). This implies that the imaginary part of the quantile spectral density* $G(\cdot)$ *is zero over all* x, y *and* ω*. In this case, we use the test statistic*

$$\mathcal{R}_T = \frac{1}{T}\sum_r \left|\Im \widehat{G}_T(x,y;\omega)\right|^2 dF(x)dF(y),$$

where F *is some distribution, and by using identical methods to those derived above we can obtain the limiting distribution of the test under the null. It is worth mentioning that Dette et al. (2011) also discuss the impact time reversibility has on the quantile spectral density.*

5 Bootstrap approximation

The asymptotic normality result that we use to obtain the p-value of the test statisic $\mathcal{Q}_T$ is only an approximation. For small samples, the normality approximation may not be particularly good, mainly because $\mathcal{Q}_T$ is a positive random variable, whose distribution will be skewed. This may well lead to more false positive than we can control for in our type I error.

To correct for this, we propose estimating the finite sample distribution of $\mathcal{Q}_T$ using a frequency domain bootstrap procedure. In a multivariate time series, the periodogram matrix at the fundamental frequencies asymptotically follow a Wishart distribution, moreover for our purposes they are close enough to be independent such that we don't loose too much information by treating them as independent (observe that the asymptotic variance of the test statistic $\mathcal{Q}_T$ is only in terms of the pair-wise distributions and does not contain any higher order dependencies). Thus motivated by the frequency domain bootstrap methods proposed in Hurvich and Zeger (1987) and Franke and Härdle (1992) for univariate data and Berkowitz and Diebold (1998) and Dette and Paparoditis (2009) for multivariate data, we propose the following bootstrap scheme to obtain an estimate of the finite sample distribution under the null hypothesis.

Let $x_1 < \cdots < x_q$ be a finite discretisation of the real line (noting that we approximate $\mathcal{Q}_T$ with the discretisation

$$\begin{aligned}\mathcal{Q}_T &= \frac{2\pi}{T}\sum_{k=1}^{T}\sum_{i_1,i_2=2}^{q}|\hat{G}(x_{i_1},x_{i_2};\omega_k) - \sum_s K_M(\omega_k-\omega_s)G_0(x_{i,1},x_{i_2};\omega_k)|^2 \times \\ &\quad (F_0(x_{i_1}) - F_0(x_{i_1-1}))(F_0(x_{i_2}) - F_0(x_{i_2-1})).\end{aligned}$$

We observe that under the null hypothesis that $\mathbf{G}_{\mathbf{Z}}(\omega)$ will be the spectral density matrix of the multivariate q-dimensional time series $\mathbf{Z}_t = (\tilde{Z}_t(x_1), \cdots, \tilde{Z}_t(x_q))$ where $\tilde{Z}_t(x) = I(X_t \le x) - F(x)$

and $\mathbf{G}_{\mathbf{Z}}(\omega)_{i_1,i_2} = G_0(x_{i_1}, x_{i_2}; \omega)$. Thus we use the transformation of X_t into a high dimensional multivariate time series to construct the the bootstrap distribution.

The steps of the frequency domain bootrap for the test statistic $\mathcal{Q}_T$ are as follows:

Step 1: Generate T independent matrices $\mathbf{I}^*_{\mathbf{Z}}(\omega_k) = \mathbf{G}_{\mathbf{Z}}(\omega_k)^{1/2} W^*_k \mathbf{G}_{\mathbf{Z}}(\omega_k)^{1/2}$, where

$$W^*_k \sim \begin{cases} W^C_q(1, I_q) & 1 \leq k \leq T/2 \\ W^R_q(1, I_q) & k \in \{0, T/2\} \\ \overline{W^*_{T-k}} & T/2 < k \leq T \end{cases},$$

and W^C and W^R denote the complex and real Wishart distributions.

Step 2: Construct the bootstrap quantile spectral density matrix estimators with $\hat{\mathbf{G}}^*_{\mathbf{Z}}(\omega_k) = \sum_s K_M(\omega_k - \omega_s)\mathbf{I}^*_{\mathbf{Z}}(\omega_s)$ for $k = 1, \ldots, T$.

Step 3: Obtain the bootstrap test statistic

$$Q^*_T = \frac{2\pi}{T} \sum_{k=1}^{T} \sum_{i_1,i_2=2}^{q} |\hat{G}^*(x_{i_1}, x_{i_2}; \omega_k) - G^M_0(x_{i_1}, x_{i_2}; \omega_k)|^2 (F_0(x_{i_1}) - F_0(x_{i_1-1}))(F_0(x_{i_2}) - F_0(x_{i_2-1})),$$

where $G^M_0(x, y; \omega) = \frac{1}{2\pi} \sum_k \lambda(\frac{k}{M}) C_{0,r}(x, y) \exp(ir\omega)$.

Step 4: Approximate the distribution of $\mathcal{Q}_T$ under the null by using the empirical distribution of the bootstrap sample $\{\mathcal{Q}^*_T\}$.

Step 5: Based on the bootstrap distribution estimate the p-value of $\mathcal{Q}_T$.

We illustrate our procedure in Figure 6, for this example we use the quantile spectral density G_0, based on an ARCH(1) ($X_t = Z_t\sigma_t$ and $\sigma^2_t = a_0 + a_1 X^2_{t-1}$), where $a_0 = 1/1.9$, $a_1 = 0.9$, Z_t are iid standard normal random variables and $T = 500$. A plot of the normal approximation, the density of $\mathcal{Q}_T$ (which is estimated and based on 500 replications) and the bootstrap estimator of the density (along with their rejection regions) is given in Figure 6. We observe that the skew in the finite sample distribution means that the normal distribution is under estimating the location of the rejection region. However, the bootstrap approximation appears to capture relatively well the finite sample distribution, and approximate well the rejection region. Since the bootstrap scheme is based on sampling from iid random variables, we can write the bootstrap test statistic as a quadratic form. Thus by using Lee and Subba Rao (2010), asymptotic normality of $\mathcal{Q}^*_T$ can be shown with mean and variance given in (4). Hence the limiting distribution of the bootstrap statistic and limiting distribution of the test statistic $\mathcal{Q}_T$, under the null, coincide.

6 Simulations and Real data examples

6.1 Simulations

In this section we conduct a simulation study. In order to determine the effectiveness of the test we will use two different models that have the same first and second order structure (thus a test based on the covariance structure would not be able to distinguish between them). In particular, we will consider the AR(1) model $X_t = \mu + aX_{t-1} + \varepsilon_t$ and the squares of the ARCH(1) model $Y_t = a_0 + aY_{t-1} + (Z_t^2 - 1)(a_0 + aY_{t-1})$, where $\{\varepsilon_t\}$ and $\{Z_t\}$ are iid zero mean Gaussian random variables with $\text{var}(Z_t) = 1$ and μ and $\text{var}(\varepsilon_t)$ chosen such that X_t and Y_t have the same mean and covariance structure. Note that in the simulation we only consider $a \leq 0.55$, so that the spectral density of the squared ARCH exists. For each model we did 1000 replications and the tests was done at both the $\alpha = 0.1$ and $\alpha = 0.05$ level.

In our simulations we used the Bartlett window, compared the test for various M and used both the normal approximation and the proposed bootstrap procedure. The results for H_0 : AR(1) against the alternative H_A : ARCH(1) (various a, fixing $a_0 = 0.4$) are given in Table 2 and 3. The results for H_0 : ARCH(1) against H_A : AR(1) are given in Table 4 and 5. We use the sample sizes $T = 100$ and 500.

As expected under the null hypothesis the null hypothesis tends to over reject, whereas the bootstrap gives a better approximation of the significance level. There appears to be very little difference in the behaviour under the null for various values of a and between the AR and the ARCH. Under the alternative, the power seems to be quite high even for quite small samples. The only model where the power is not close to 100% is when $a = 0.3$, sample size $T = 100$, the null is an AR(1) and the alternative is an ARCH(1). This can be explained by the fact that for small values of a, both the AR and the ARCH models are relatively close to independent observations, thus making it relatively difficult to reject the null.

6.2 Real Data

In this section we consider the the Microsoft daily return data (March, 1986 - June, 2003) discussed in Section 2.1 and the Intel monthly log return data (January 1973 - December 2003). In the analysis below we will test whether the GARCH and ARCH models are appropriate for the Microsoft and Intel data, respectively. We use the Bartlett window.

A plot of the estimated $\widehat{G}_T$ together with the piece-wise confidence intervals (obtained using the results in Theorem 3.1) and the corresponding quantile spectral density of the GARCH(1,1) is given in Figure 7 for the Microsoft data. It is clear from the plot that the GARCH$(1,1)$ model with coefficients evaluated using the maximum likelihood estimator is not the appropriate model

to fit to this data. The plots suggest that the main deviation from the GARCH$(1,1)$ arises at about $x, y = 0$, This observation is confirmed by the results of our test. Using various values of M ranging from $30 - 70$, the p-value corresponding to $\mathcal{Q}_T$ is almost zero with both the normal approximation and also the Bootstrap method. Therefore, from our analysis it seems that the GARCH$(1,1)$ is not a suitable model for modelling the Microsoft daily returns from 1986-2003.

We now consider the second data set, the Intel monthly log returns from 1973 - 2003. Tsay (2005) propose fitting an ARCH(1) (with Gaussian innovations) model to this data, and maximum likelihood yields the estimators $\mu = 0.0166$, $a_0 = 0.0125$ and $a_1 = 0.363$, where $X_t = \mu + \varepsilon_t$, $\varepsilon_t = \sigma_t Z_t$ and $\sigma_t^2 = a_0 + a_1 \varepsilon_{t-1}^2$. A plot of the estimated $\widehat{G}_T$ with the piece-wise confidence intervals together the quantile spectral density of the ARCH(1) model is given in Figure 8. We observe that quantile spectral density of the ARCH model lies in the confidence intervals for almost all frequencies. These observations are confirmed by the proposed goodness of fit test. A summary of the results for various M, using both the normal approximation and the bootstrap method is given in Table 1. The p-values for the normal approximation tend to be smaller than the p-values of the bootstrap method, this is probably due to the skew in the finite sample distribution which results in smaller p-values. However, both the normal approximation and the bootstrap give relatively large p-values for all values of M. Therefore there is not enough evidence to reject the null. This backs the claims in Tsay (2005) that the ARCH(1) may be an appropriate model for the the Intel data.

M	15	20	25	30
Normal p-value	0.0905	0.1279	0.1807	0.2643
Bootstrap p-value	0.3880	0.4320	0.4020	0.4780

Table 1: The p-values for the Intel Data and various values of M

Acknowledgments

This work has been partially supported by the National Science Foundation, DMS-1106518.

A Proofs

To obtain the sampling properties of $\hat{G}_T(\cdot)$ and $\mathcal{Q}_T$ (under both the null and alternative), we first replace the empirical distribution function $\hat{F}_T(x)$, with the true distribution and show that the error is negligible. Define the zero mean, transformed variable $\tilde{Z}_t(x) = I(X_t \leq x) - F(x)$, where $F(\cdot)$ denotes the marginal distribution of $\{X_t\}$. In addition define $\tilde{C}_r(x,y) = \frac{1}{T}\sum_t \tilde{Z}_t(x)\tilde{Z}_{t+r}(y)$,

$$\tilde{G}_T(x,y;\omega_k) = \frac{1}{2\pi}\sum_r \lambda_M(r)\tilde{C}_r(x,y)\exp(ir\omega_k) = \sum_s K_M(\omega_k-\omega_s)\tilde{J}_T(x;\omega_s)\overline{\tilde{J}_T(y;\omega_s)},$$

$$\tilde{\mathcal{Q}}_T = \frac{1}{T}\sum_{k=1}^{T}\int |\tilde{G}_T(x,y;\omega_k) - \sum_r \lambda_M(r)C_{0,r}(x,y)\exp(ir\omega_k)|^2 dF_0(x)dF_0(y).$$

where $\tilde{J}_T(x;\omega) = \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^T Z_t(x)\exp(it\omega)$.

In the proofs below we shall use the notation $\|X\|_r = (\mathbb{E}(|X|^r))^{1/r}$. We first show that replacing $\hat{F}_T(x)$ with $F(x)$ does not affect the asymptotic sampling properties of $G_T(\cdot)$ and $\mathcal{Q}_T$.

Lemma A.1 *Suppose Assumption 3.1 holds. Then we have*

$$(\mathbb{E}|\widehat{G}_T(x,y;\omega) - \tilde{G}_T(x,y;\omega)\big|^2)^{1/2} = O(\frac{M}{T}) \tag{8}$$

and

$$(\mathbb{E}|\mathcal{Q}_T - \tilde{\mathcal{Q}}_T\big|^2)^{1/2} = O(\frac{1}{T}). \tag{9}$$

PROOF. We first observe that

$$\begin{aligned} & J_T(x;\omega_k)\overline{J_T(y;\omega_k)} - \tilde{J}_T(x;\omega_k)\overline{\tilde{J}_T(y;\omega_k)} \\ = & \begin{cases} 0 & \omega_k \neq 0,\pi \\ T(\hat{F}_T(x)-F(x))(\hat{F}_T(y)-F(y)) & \text{otherwise} \end{cases}. \end{aligned}$$

Substituting the above into $\hat{G}_T(\omega_s) - \tilde{G}_T(\omega_s)$ gives

$$\hat{G}_T(\omega_s) - \tilde{G}_T(\omega_s) = TK_M(\omega_s)(\hat{F}_T(x)-F(x))(\hat{F}_T(y)-F(y)). \tag{10}$$

Using $K_M(\cdot) = O(\frac{M}{T})$ and $||\hat{F}_T(x) - F(x)||_2 = O(\frac{1}{T})$ in (10), we obtain the desired result for (8).

To prove (9) note that

$$
\begin{aligned}
& \mathcal{Q}_T - \tilde{\mathcal{Q}}_T \\
= & \int \frac{1}{T}\sum_{s=1}^{T} \big(\hat{G}_T(x,y;\omega_s) - \tilde{G}_T(x,y;\omega_s)\big)\overline{\big(\hat{G}_T(x,y;\omega_s) + \tilde{G}_T(x,y;\omega_s)\big)} dF_0(x)dF_0(y) \\
+ & \Re\bigg(\int \frac{2}{T}\sum_{s=1}^{T} \big(\hat{G}_T(x,y;\omega_s) - \tilde{G}_T(x,y;\omega_s)\big)G(x,y;\omega_s)dF_0(x)dF_0(y)\bigg).
\end{aligned}
$$

Thus substituting (10) into the above gives

$$
\begin{aligned}
& \mathcal{Q}_T - \tilde{\mathcal{Q}}_T \\
= & \int (\widehat{F}_T(x) - F(x))(\widehat{F}_T(y) - F(y)) \times \\
& \bigg(\sum_{s=1}^{T} K_M(\omega_s)\big(\hat{G}_T(x,y;\omega_s) + \tilde{G}_T(x,y;\omega_s)\big)\bigg) dF_0(x)dF_0(y) \\
+ & 2\int (\widehat{F}_T(x) - F(x))(\widehat{F}_T(y) - F(y))\Re\bigg(\sum_{s=1}^{T} K_M(\omega_s)G(x,y;\omega_s)\bigg)dF_0(x)dF_0(y).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& \big\|\mathcal{Q}_T - \tilde{\mathcal{Q}}_T\big\|_2 \\
\leq & \int \big\|\widehat{F}_T(x) - F(x)\big\|_8 \big\|\widehat{F}_T(y) - F(y)\big\|_8 \times \\
& \bigg(\sum_{s=1}^{T} \big(\big|K_M(\omega_s)\big| \cdot \big(\big\|\hat{G}_T(x,y;\omega_s)\big\|_8 + \|\tilde{G}_T(x,y;\omega_s)\|_8\big)\big)\bigg) dF_0(x)dF_0(y) \\
+ & 2\int \big\|\widehat{F}_T(x) - F(x)\big\|_4 \big\|\widehat{F}_T(y) - F(y)\big\|_4 \times \\
& \bigg(\sum_{s=1}^{T} \big|K_M(\omega_s)\big| \cdot \big|G(x,y;\omega_s)\big|\bigg) dF_0(x)dF_0(y).
\end{aligned}
$$

For all $r \geq 2$, we have $\|\widehat{F}_T(x) - F(x)\|_r = O(\frac{1}{\sqrt{T}})$, substituting this into the above gives $\big\|\mathcal{Q}_T - \tilde{\mathcal{Q}}_T\big\|_2 = O(\frac{1}{T})$, and the desired result. □

PROOF of Theorem 3.1 To show asymptotic normality of $\widehat{G}_T(\cdot)$, we first replace $\widehat{G}_T$ with $\tilde{G}_T$, by (8) the replacement error is $O_p(\frac{M}{T})$. Thus $\widehat{G}_T$ and $\tilde{G}_T$ have the same asymptotic distribution and we can show how asymptotic normality of $\widehat{G}_T$ by considering $\tilde{G}_T(\cdot)$ instead. To show asymptotic normality of $\tilde{G}_T$ we use identical methods to those in Lee and Subba Rao (2011), where, since $\{I(X_t < x)\}$ are bounded random variables, we can use Ibragimov's covariance bounds for bounded random variables. To obtain the limiting variance we note that under

Assumption 3.1, since $s > 2$, we have that $\sum_r |r| \cdot |\mathrm{cov}(I(X_0 \leq x), I(X_r \leq y))| < \infty$ and $\sum_{r_1,r_2,r_2}(1+|r_j|)|\mathrm{cum}(I(X_0 \leq x_0), I(X_{r_1} \leq x_1), I(X_{r_2} \leq x_2), I(X_{r_3} \leq x_3))| < \infty$. Thus, the assumptions in Brillinger (1981), Theorem 3.4.3 are satisfied, which allows us to obtain the stated limiting variance. $\square$

We use the following lemma to obtain a bound for the variance of $\mathcal{Q}_T$.

Lemma A.2 *Let the lag window be defined as in Definition 2.1 and suppose $h_1(\cdot)$ and $h_2(\cdot)$ are bounded functions. Then we have*

$$L_1 = \int h_1(u_1)h_2(u_2)\Delta_M(u_1-u_2)^2 du_1 du_2 = O(M) \tag{11}$$

and

$$L_2 = \int h_1(u_1)h_2(u_2)\Delta_M(u_1+u_2)\Delta_M(u_1-u_2) du_1 du_2 = O(1) \tag{12}$$

where $\Delta_M(\cdot)$ is defined in (5).

PROOF. To simplify notation we prove the result for the truncated lag window $\lambda(u) = I_{[-1,1]}(u)$, but a similar result can also be proven for lag windows which satisfy Definition 2.1. In the proof we use the following two identities

$$\sum_{t=0}^{T} e^{it\omega} = e^{\frac{iT\omega}{2}}\frac{\sin(\frac{T+1}{2}\omega)}{\sin(\omega/2)} \quad \text{and} \quad \Big(\int \Big|\frac{\sin(\frac{M+1}{2}(u))}{\sin((u)/2)}\Big|^p du\Big)^{1/p} = O(M^{1-p^{-1}}). \tag{13}$$

We start by expanding Δ_M and using the above, to give

$$\begin{aligned}
\Delta_M(\theta_1-\theta_2) &= \int \sum_{j_1,j_2=-M}^{M} \lambda_M(j_1)\lambda_M(j_2)\exp(ij_1(\omega_{s_1}-\theta))\exp(ij_2(\omega_{s_2}-\theta))d\omega \\
&= \sum_j \lambda_M(j)\lambda_M(-j)\exp(ij(\theta_1-\theta_2)) \\
&= \frac{\sin((M+1)(\theta_1-\theta_2)/2)}{\sin((\theta_1-\theta_2)/2)} 2\Re e^{\frac{iM(\theta_1-\theta_2)}{2}}.
\end{aligned} \tag{14}$$

Substituting the above and (13) into (11) gives

$$|L_1| = \Big|\int\int h_1(u_1)h_2(u_2)\Delta_M(u_1-u_2)^2 du_1 du_2\Big| \leq \sup_{u,i}|h_i(u)|^2 \int\int \Big|\frac{\sin(\frac{M+1}{2}(u_1-u_2))}{\sin((u_1-u_2)/2)}\Big|^2 du_1 du_2 = O(M).$$

This proves (11). To prove (12) we observe that by a change of variables ($v_1 = u_1 - u_2$ and

$v_2 = u_1 + u_2$) we have

$$|L_2| \;\leq\; C\int |\Delta_M(u_1+u_2)|\cdot|\Delta_M(u_1-u_2)|du_1du_2 \leq C\Big(\int |\Delta_M(u)|du\Big)^2.$$

Now by substituting (14) and (13) into the above gives $L_2 = O(1)$. Thus we have obtained the desired result. □

PROOF Lemma 3.1 We first evalulate the expectation of $\mathcal{Q}_T$. By using Lemma A.1 we have

$$\begin{aligned}
&\mathbb{E}(\mathcal{Q}_T)\\
&= \frac{1}{T}\sum_{s=1}^{T}\int\sum_{k_1,k_2=1}^{T} K_M(\omega_s-\omega_{k_1})K_M(\omega_s-\omega_{k_2})\mathrm{cov}\big(\tilde{J}_T(x;\omega_{k_1})\overline{\tilde{J}_T(y;\omega_{k_1})},\tilde{J}_T(x;\omega_{k_2})\overline{\tilde{J}_T(y;\omega_{k_2})}\big)\big) + O(\frac{1}{T})\\
&= I_1 + I_2 + I_3 + O(\frac{1}{T}),
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \frac{1}{T}\int\sum_{s,k_1,k_2=1}^{T}\prod_{i=1}^{2}K_M(\omega_s-\omega_{k_i})\mathrm{cov}\big(\tilde{J}_T(x;\omega_{k_1}),\tilde{J}_T(x;\omega_{k_2}))\mathrm{cov}(\overline{\tilde{J}_T(y;\omega_{k_2})},\overline{\tilde{J}_T(y;\omega_{k_2})})dF_0(x)dF_0(y)\\
I_2 &= \frac{1}{T}\int\sum_{s,k_1,k_2=1}^{T}\prod_{i=1}^{2}K_M(\omega_s-\omega_{k_i})\mathrm{cov}\big(\tilde{J}_T(x;\omega_{k_1}),\overline{\tilde{J}_T(y;\omega_{k_2})})\mathrm{cov}(\overline{\tilde{J}_T(y;\omega_{k_1})},\tilde{J}_T(y;\omega_{k_2}))dF_0(x)dF_0(y)\\
I_3 &= \frac{1}{T}\int\sum_{s,k_1,k_2=1}^{T}\prod_{i=1}^{2}K_M(\omega_s-\omega_{k_i})\mathrm{cum}\big(\tilde{J}_T(x;\omega_{k_1}),\overline{J}_T(y;\omega_{k_1}),\tilde{J}_T(x;\omega_{k_2}),\overline{\tilde{J}}_T(y;\omega_{k_2})\big)dF_0(x)dF_0(y).
\end{aligned}$$

Under Assumption 3.1, we have that $\sum_r |r|\cdot|\mathrm{cov}(I(X_0\leq x),I(X_r\leq y))| < \infty$ and $\sum_{r_1,r_2,r_2}(1+|r_j|)|\mathrm{cum}(I(X_0\leq x_0),I(X_{r_1}\leq x_1),I(X_{r_2}\leq x_2),I(X_{r_3}\leq x_3))| < \infty$. Therefore we can apply Brillinger (1981), Theorem 3.4.3 to obtain

$$\begin{aligned}
I_1 &= \frac{1}{T}\sum_{s=1}^{T}\int\sum_{k=1}^{T}K_M(\omega_s-\omega_k)^2\int G(x,x;\omega_k)G(y,y;\omega_k)dF_0(x)dF_0(y) + O(\frac{1}{T}) = O(\frac{M}{T})\\
I_2 &= \frac{1}{T}\sum_{s=1}^{T}\int\sum_{k=1}^{T}K_M(\omega_s-\omega_k)K_M(\omega_s+\omega_k)\int G(x,y;\omega_k)G(y,x;\omega_k)dF_0(x)dF_0(y) + O(\frac{1}{T}) = O(\frac{1}{T})\\
I_3 &= \frac{1}{T^2}\int\sum_{r}\lambda_M(r)^2\sum_{t_1,t_2=1}^{T}\mathrm{cum}(Z_{t_1}(x),Z_{t_1+r}(y),Z_{t_2}(x),Z_{t_2+r}(y))dF_0(x)dF_0(y) = O(\frac{1}{T}).
\end{aligned}$$

This gives us an asymptotic expression for the expectation. We now obtain an expression for the

variance. Replacing $Z_t(\cdot)$ with $\tilde{Z}_t(\cdot)$ gives

$$
\begin{aligned}
&\mathrm{var}(\mathcal{Q}_T) = \\
&\frac{1}{T^2}\sum_{s_1,s_2=1}^{T}\int\Big(\sum_{k_1,k_2,k_3,k_4} K_M(\omega_{s_1}-\omega_{k_1})K_M(\omega_{s_1}-\omega_{k_2})K_M(\omega_{s_2}-\omega_{k_3})K_M(\omega_{s_2}-\omega_{k_4}) \\
&\times\mathrm{cov}\big((J_{k_1,x_1}\overline{J}_{k_1,y_1}-\mathbb{E}(J_{k_1,x_1}\overline{J}_{k_1,y_1}))(J_{k_2,x_1}\overline{J}_{k_2,y_1}-\mathbb{E}(J_{k_2,x_1}\overline{J}_{k_2,y_1})), \\
&(J_{k_3,x_2}\overline{J}_{k_3,y_2}-\mathbb{E}(J_{k_3,x_2}\overline{J}_{k_3,y_2}))(J_{k_4,x_2}\overline{J}_{k_4,y_2}-\mathbb{E}(J_{k_4,x_2}\overline{J}_{k_4,y_2}))\big)\Big)dF_0(x_1)dF_0(y_1)dF_0(x_2)dF_0(y_2)+O(\frac{1}{T}) \\
&= II_1+II_2+II_3+O(\frac{1}{T})
\end{aligned}
$$

where $J_{k,x}=\tilde{J}_T(x;\omega_k)$,

$$
\begin{aligned}
II_1 &= \frac{1}{T^2}\sum_{s_1,s_2=1}^{T}\int\sum_{k_1,k_2,k_3,k_4}\mathrm{cum}(J_{k_1,x_1}\bar{J}_{k_1,y_1},\bar{J}_{k_3,x_2}J_{k_3,y_2})\mathrm{cum}(J_{k_2,x_1}\bar{J}_{k_2,y_1},\overline{J}_{k_4,x_2}J_{k_4,y_2}) \\
&\quad \prod_{i=1}^{2}K_M(\omega_{s_1}-\omega_{k_i})\prod_{i=3}^{4}K_M(\omega_{s_2}-\omega_{k_i})dF_0(x_1)dF_0(y_1)dF_0(x_2)dF_0(y_2) \\
II_2 &= \frac{1}{T^2}\sum_{s_1,s_2=1}^{T}\int\sum_{k_1,k_2,k_3,k_4}\mathrm{cum}(J_{k_1,x_1}\bar{J}_{k_1,y_1},\overline{J}_{k_4,x_2}J_{k_4,y_2})\mathrm{cum}(J_{k_2,x_1}\bar{J}_{k_2,y_1},\bar{J}_{k_3,x_2}J_{k_3,y_2}) \\
&\quad \prod_{i=1}^{2}K_M(\omega_{s_1}-\omega_{k_i})\prod_{i=3}^{4}K_M(\omega_{s_2}-\omega_{k_i})dF_0(x_1)dF_0(y_1)dF_0(x_2)dF_0(y_2) \\
II_3 &= \frac{1}{T^2}\sum_{s_1,s_2=1}^{T}\int\sum_{k_1,k_2,k_3,k_4}\mathrm{cum}(J_{k_1,x_1}\bar{J}_{k_1,y_1},J_{k_2,x_1}\bar{J}_{k_2,y_1},\bar{J}_{k_3,x_2}J_{k_3,y_2},\overline{J}_{k_4,x_2}J_{k_4,y_2}) \\
&\quad \prod_{i=1}^{2}K_M(\omega_{s_1}-\omega_{k_i})\prod_{i=3}^{4}K_M(\omega_{s_2}-\omega_{k_i})dF_0(x_1)dF_0(y_1)dF_0(x_2)dF_0(y_2).
\end{aligned}
$$

To obtain an expression for the variance we start by expanding II_1

$$
\begin{aligned}
II_1 \;&= \frac{1}{T^2}\sum_{s_1,s_2}\int\sum_{k_1,k_2,k_3,k_4}\prod_{i=1}^{2}K_M(\omega_{s_1}-\omega_{k_i})\prod_{i=3}^{4}K_M(\omega_{s_2}-\omega_{k_i}) \\
\times &\Big(\mathrm{cov}(J_{k_1,x_1},J_{k_3,x_2})\mathrm{cov}(\overline{J}_{k_1,y_1},\overline{J}_{k_3,y_2})\mathrm{cov}(J_{k_2,x_1},J_{k_4,x_2})\mathrm{cov}(\overline{J}_{k_2,y_1},\overline{J}_{k_4,y_2}) \\
+ &\quad \mathrm{cov}(J_{k_1,x_1},J_{k_3,x_2})\mathrm{cov}(\overline{J}_{k_1,y_1},\overline{J}_{k_3,y_2})\mathrm{cov}(J_{k_2,x_1},\overline{J}_{k_4,y_2})\mathrm{cov}(\overline{J}_{k_2,y_1},J_{k_4,x_2}) \\
+ &\quad \mathrm{cov}(J_{k_1,x_1},J_{k_3,x_2})\mathrm{cov}(\overline{J}_{k_1,y_1},\overline{J}_{k_3,y_2})\mathrm{cum}(J_{k_2,x_1},\overline{J}_{k_2,y_1},J_{k_4,x_2},\overline{J}_{k_4,y_2}) \\
+ &\quad \mathrm{cov}(J_{k_1,x_1},\overline{J}_{k_3,y_2})\mathrm{cov}(\overline{J}_{k_1,y_1},J_{k_3,x_2})\mathrm{cov}(J_{k_2,x_1},J_{k_4,x_2})\mathrm{cov}(\overline{J}_{k_2,y_1},\overline{J}_{k_4,y_2}) \\
+ &\quad \mathrm{cov}(J_{k_1,x_1},\overline{J}_{k_3,y_2})\mathrm{cov}(\overline{J}_{k_1,y_1},J_{k_3,x_2})\mathrm{cov}(J_{k_2,x_1},\overline{J}_{k_4,y_2})\mathrm{cov}(\overline{J}_{k_2,y_1},J_{k_4,x_2}) \\
+ &\quad \mathrm{cov}(J_{k_1,x_1},\overline{J}_{k_3,y_2})\mathrm{cov}(\overline{J}_{k_1,y_1},J_{k_3,x_2})\mathrm{cum}(J_{k_2,x_1},\overline{J}_{k_2,y_1},J_{k_4,x_2},\overline{J}_{k_4,y_2}) \\
+ &\quad \mathrm{cum}(J_{k_1,x_1},\overline{J}_{k_1,y_1},J_{k_3,x_2},\overline{J}_{k_3,y_2})\mathrm{cov}(J_{k_2,x_1},J_{k_4,x_2})\mathrm{cov}(\overline{J}_{k_2,y_1},\overline{J}_{k_4,y_2}) \\
+ &\quad \mathrm{cum}(J_{k_1,x_1},\overline{J}_{k_1,y_1},J_{k_3,x_2},\overline{J}_{k_3,y_2})\mathrm{cov}(J_{k_2,x_1},\overline{J}_{k_4,y_2})\mathrm{cov}(\overline{J}_{k_2,y_1},J_{k_4,x_2}) \\
+ &\quad \mathrm{cum}(J_{k_1,x_1},\overline{J}_{k_1,y_1},J_{k_3,x_2},\overline{J}_{k_3,y_2})\mathrm{cum}(J_{k_2,x_1},\overline{J}_{k_2,y_1},J_{k_4,x_2},\overline{J}_{k_4,y_2})\Big)\prod_{j=1}^{2}dF_0(x_j)dF_0(y_j) \\
:= &\quad \sum_{j=1}^{9} II_{1,j}.
\end{aligned}
$$

We use (Brillinger, 1981), Theorem 3.4.3 to obtain the following expression for $II_{1,1}$

$$
\begin{aligned}
II_{1,1} = \;& \frac{1}{T^2}\sum_{s_1,s_2}\int\Big(\sum_{k_1,k_2=1}^{T}\big(\prod_{i=1}^{2}K_M(\omega_{s_i}-\omega_{k_1})K_M(\omega_{s_i}-\omega_{k_2})\big) \\
& \mathrm{cov}(J_{k_1,x_1},J_{k_1,x_2})\mathrm{cov}(\overline{J}_{k_1,y_1},\overline{J}_{k_1,y_2})\mathrm{cov}(J_{k_2,x_1},J_{k_2,x_2})\mathrm{cov}(\overline{J}_{k_2,y_1},\overline{J}_{k_2,y_2})\Big) \\
& \prod_{j=1}^{2}dF_0(x_j)dF_0(y_j)+O(\frac{1}{T^2}) \\
= & \frac{1}{T^2}\int\int\Big(\int W_M(\omega_{s_1}-\theta_1)W_M(\omega_{s_1}-\theta_2)d\omega_{s_1}\Big)\times \\
& \Big(\int W_M(\omega_{s_2}-\theta_1)W_M(\omega_{s_2}-\theta_2)d\omega_{s_2}\Big)\prod_{i=1}^{2}G(x_1,x_2;\theta_i)G(y_1,y_2;-\theta_i)d\theta_i \\
& \prod_{j=1}^{2}dF_0(x_j)dF_0(y_j)+O(\frac{1}{T^2}) \\
= & \frac{1}{T^2}\int\int\Delta_M(\theta_1-\theta_2)^2\prod_{i=1}^{2}G(x_1,x_2;\theta_i)G(y_1,y_2;-\theta_i)d\theta_i\prod_{j=1}^{2}dF_0(x_j)dF_0(y_j)+O(\frac{1}{T^2}).
\end{aligned}
$$

Therefore by using (11) we have $II_{1,1}=O(\frac{M}{T^2})$. We now consider $II_{1,2}$, by using a similar method

we have

$$
\begin{aligned}
II_{1,2} &= \frac{1}{T^2}\int \Big(W_M(\omega_{s_1}-\theta_1)W_M(\omega_{s_1}-\theta_2)W_M(\omega_{s_2}-\theta_1)W_M(\omega_{s_2}+\theta_2)\times \\
&\quad G(x_1,x_2,\theta_1)G(y_1,y_2,-\theta_1)G(x_1,y_2,\theta_2)G(y_1,x_2,-\theta_2)\Big)d\theta_1 d\theta_2 d\omega_{s_1}d\omega_{s_2}\prod_{j=1}^{2}dF_0(x_j)dF_0(y_j)+O(\frac{1}{T^2}) \\
&= \frac{1}{T^2}\int \Delta_M(\theta_1-\theta_2)\Delta_M(\theta_1+\theta_2)G(x_1,x_2,\theta_1)G(y_1,y_2,-\theta_1) \\
&\quad G(x_1,y_2,\theta_2)G(y_1,x_2,-\theta_2)d\theta_1 d\theta_2\prod_{j=1}^{2}dF_0(x_j)dF_0(y_j)+O(\frac{1}{T^2}).
\end{aligned}
$$

By using (12) the above integral is $O(1)$, and altogether $II_{1,2}=O(\frac{1}{T^2})$. Using a similar argument, one can show that $II_{1,3}$, $II_{1,4}$ are smaller than $O(\frac{M}{T^2})$, so negligible. For $II_{1,5}$, we use that

$$
\mathrm{cov}(J_{k_1,x},\overline{J}_{k_2,y}) = \begin{cases} G(x,y,\omega_{k_1}) & k_1+k_2=T \\ O(\frac{1}{T}) & otherwise \end{cases}
$$

, which follows from (Brillinger, 1981), Theorem 3.4.3. This leads to

$$
\begin{aligned}
II_{1,5} &= \frac{1}{T^2}\sum_{s_1,s_2}\int\Big(\sum_{k_1,k_2=1}^{T}(\prod_{i=1}^{2}K_M(\omega_{s_1}-\omega_{k_i})K_M(\omega_{s_2}+\omega_{k_i})) \\
&\quad \mathrm{cov}(J_{k_1,x_1},J_{k_1,y_2})\mathrm{cov}(\overline{J}_{k_1,y_1},\overline{J}_{k_1,x_2})\mathrm{cov}(J_{k_2,x_1},J_{k_2,y_2})\mathrm{cov}(\overline{J}_{k_2,y_1},\overline{J}_{k_2,x_2})\Big) \\
&\quad \prod_{j=1}^{2}dF_0(x_j)dF_0(y_j)+O(\frac{1}{T^2}) \\
&= \frac{1}{T^2}\int\int\Big(\int W_M(\omega_{s_1}-\theta_1)W_M(\omega_{s_1}-\theta_2)d\omega_{s_1}\Big)\times \\
&\quad \Big(\int W_M(\omega_{s_2}+\theta_1)W_M(\omega_{s_2}+\theta_2)d\omega_{s_2}\Big)\prod_{i=1}^{2}G(x_1,y_2;\theta_i)G(y_1,x_2;-\theta_i)d\theta_i \\
&\quad \prod_{j=1}^{2}dF_0(x_j)dF_0(y_j)+O(\frac{1}{T^2}) \\
&= II_{1,1}
\end{aligned}
$$

because of $\Delta(\theta)=\Delta(-\theta)$ and interchangeability of integrals about (x_1,x_2,y_1,y_2). With a similar method, one can show that $II_{1,6}\ldots,II_{1,9}$ are all dominated by $II_{1,1}$ and $II_{1,5}$ Altogether this

gives

$$II_1 = \frac{2}{T^2} \int \Delta_M(\theta_1 - \theta_2)^2 \prod_{i=1}^{2} G(x_1, x_2; \theta_i) G(y_1, y_2; \theta_i) d\theta_i \prod_{j=1}^{2} dF_0(x_j) dF_0(y_j) + O(\frac{1}{T^2}).$$

Using the identical argument with the above, we can show that

$$II_2 = \frac{2}{T^2} \int \Delta_M(\theta_1 - \theta_2)^2 \prod_{i=1}^{2} G(x_1, x_2; \theta_i) G(y_1, y_2; \theta_i) d\theta_i \prod_{j=1}^{2} dF_0(x_j) dF_0(y_j) + O(\frac{1}{T^2}).$$

To bound II_3 we recall that

$$\begin{aligned} II_3 &= \frac{1}{T^2} \sum_{s_1, s_2=1}^{T} \int \sum_{k_1, k_2, k_3, k_4} K_M(\omega_{s_1} - \omega_{k_1}) K_M(\omega_{s_1} - \omega_{k_2}) K_M(\omega_{s_2} - \omega_{k_3}) K_M(\omega_{s_2} - \omega_{k_4}) \\ &\quad \mathrm{cum}\big(J_{k_1, x_1} \overline{J}_{k_1, y_1}, J_{k_2, x_1} \overline{J}_{k_2, y_1}, J_{k_3, x_2} \overline{J}_{k_3, y_2}, J_{k_4, x_2} \overline{J}_{k_4, y_2}\big) dF_0(x_1) dF_0(y_1) dF_0(x_2) dF_0(y_2) + O(\frac{1}{T}). \end{aligned}$$

By using the method of indecomposable partitions (see (Brillinger, 1981), Theorem 2.3.2) to partition the above cumulant of products into the product of cumulants. This together with (Brillinger, 1981), Theorem 3.4.3 gives us $II_3 = O(\frac{M}{T^3})$ (see Lee's thesis for further details).

Combining the expressions for II_1, II_2 and II_3 gives us the expression for the variance and completes the proof. □

A.1 Proof of Theorem 3.2

Now we show that $\mathcal{Q}_T$ can be approximated by the sum of martingale differences, this will allow us to use the martingale central limit theorem to prove Theorem 3.2.

We first define the martingale difference decomposition of $\tilde{Z}_t(x) = \sum_{j=0}^{t} M_j^{(x)}(t-j)$, where $M_j^{(x)}(t-j) = \mathbb{E}(Z_t(x)|\mathcal{F}_{t-j}) - \mathbb{E}(Z_t(x)|\mathcal{F}_{t-j-1})$, where for $t > 0$ we have $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots, X_1)$ and for $t \le 0$ we let $\mathcal{F}_t = \sigma(1)$, and $M_j(s) = 0$ for $j \ge s$. Using the above notation we define the random variable

$$\begin{aligned} \mathcal{S}_T &= \frac{1}{T^2} \int \sum_{j_1, \ldots, j_4 = 0}^{\infty} \sum_{t_1, r, t_2 \in \mathcal{A}} \lambda_M(r)^2 M_{j_1}^{(x)}(t_1 - j_1) M_{j_2}^{(y)}(t_1 + r - j_2) \\ &\quad \times M_{j_3}^{(x)}(t_2 - j_3) M_{j_4}^{(y)}(t_2 + r - j_4) dF_0(x) dF_0(y), \end{aligned} \qquad (15)$$

where $\mathcal{A} = \{(t_1 - j_1, t_1 + r - j_2, t_2 - j_3, t_2 + r - j_4)$ are all different $\}$.

Theorem A.1 *Suppose Assumption 3.1 holds, $\mathcal{S}_T$ is defined as in (15) and the null hypothesis*

is true. Then we have

$$\mathcal{Q}_T - \mathbb{E}(\mathcal{Q}_T) \;=\; \mathcal{S}_T + O_p(\frac{1}{T} + \frac{M^{1/2}}{T^{3/2}})$$

and for all $r \geq 2$ $\left\|\mathcal{S}_T\right\|_r = O(\frac{M^{1/2}}{T})$.

PROOF. We prove the result using a combination of iterative martingales and Burkholder's inequality for martingale differences. We note that for $r \geq 2$ we have

$$\|M_j^{(x)}(t-j)\|_r \;=\; \|\mathbb{E}(Z_t(x)|\mathcal{F}_{t-j}) - \mathbb{E}(Z_t(x)|\mathcal{F}_{t-j-1})\|_2 \leq 2\|\mathbb{E}(Z_t(x)|\mathcal{F}_{t-j})\|_r \leq C\alpha(j), \quad (16)$$

where $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots, X_1)$, which follows from Ibragimov's inequality. Substituting the representation $Z_t(x) = \sum_{j=1}^{\infty} M_j^{(x)}(t-j)$ into $\mathcal{Q}_T$ gives

$$\begin{aligned}
&\mathcal{Q}_T - \mathbb{E}(\mathcal{Q}_T) \\
= \;& \frac{1}{T^2}\sum_{j_1,\ldots,j_4=0}^{\infty}\sum_{r=-M}^{M}\lambda_M(r)^2\sum_{t_1,t_2}\Big(\overline{M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1+r-j_2)} \times \overline{M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4)} \\
- \;& \mathbb{E}\big(\overline{M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1+r-j_2)} \times \overline{M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4)})\big)\Big)dF_0(x)dF_0(y),
\end{aligned}$$

where $\overline{X}$ denotes the centralised random variable $X - \mathbb{E}(X)$ (note that $M_j(s) = 0$ for $s \leq 0$). We now partition the above sum into several cases, where we treat $j_1, \ldots, j_4$ as free and condition on t_1, t_2 and r:

(i) $\mathcal{A} = \{(t_1, t_2, r)$ such that $(t_1 - j_1, t_1 + r - j_2, t_2 - j_3, t_2 + r - j_4)$ are all different $\}$.

(ii) $\mathcal{B} = \{(t_1, t_2, r)$ such that $(t_1 - j_1 = t_1 + r - j_2)$ and $(t_2 - j_3 = t_2 + r - j_4)\}$.

(iii) $\mathcal{C} = \{(t_1, t_2, r)$ such that $(t_1 - j_1) = (t_2 - j_3)$ or $(t_2 + r - j_4)$ and $(t_1 + r - j_2) \neq (t_1 - j_1)\}$.

(iv) $\mathcal{D} = \{(t_1, t_2, r)$ such that $(t_1 + r - j_2) = (t_2 - j_3)$ or $(t_2 + r - j_4)$ and $(t_1 - j_1) \neq (t_1 + r - j_2)\}$.

(v) $\mathcal{E} = \{(t_1, t_2, r)$ such that $(t_2 - j_3) = (t_1 - j_1)$ or $(t_1 + r - j_2)$ and $(t_2 + r - j_4 \neq t_2 - j_3)\}$.

(iv) $\mathcal{F} = \{(t_1, t_2, r)$ such that $(t_2 + r - j_4) = (t_1 - j_2)$ or $(t_1 + r - j_2)$ and $(t_2 - j_3) \neq (t_2 + r - j_4)\}$.

Thus

$$\mathcal{Q}_T - \mathbb{E}(\mathcal{Q}_T) = \int \Big(I_{\mathcal{A}} + I_{\mathcal{B}} + I_{\mathcal{C}} + I_{\mathcal{D}} + I_{\mathcal{E}} + I_{\mathcal{F}}\Big)dF_0(x)dF_0(y),$$

where

$$I_{\mathcal{A}} = \frac{1}{T^2}\sum_{j_1,\ldots,j_4=0}^{\infty}\sum_{r,t_1,t_2\in\mathcal{A}}\lambda_M(r)^2 M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1+r-j_2)M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4),$$

$$\begin{aligned} I_{\mathcal{B}} &= \frac{1}{T^2}\sum_{j_1,\ldots,j_4=0}^{\infty}\sum_{r,t_1,t_2\in\mathcal{B}}\lambda_M(j_1-j_2)^2\Big(\overline{M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1-j_1)}\times\overline{M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2-j_3)}- \\ &\quad \mathbb{E}\big(\overline{M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1-j_1)}\times\overline{M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2-j_3)}\big)\Big)\end{aligned}$$

for $I_{\mathcal{C}},\ldots,I_{\mathcal{F}}$ are defined similarly.

We first bound $I_{\mathcal{A}}$. We partition $\mathcal{A}$ into 24 cases by the order of $(t_1-j_1,t_1+r-j_2,t_2-j_3,t_2+r-j_4)$. The first is $\mathcal{A}_1=\{(t_1,t_2,r)$ such that $t_1-j_1>t_1+r-j_2>t_2-j_3>t_2+r-j_4\}$ which gives

$$I_{\mathcal{A},1} = \frac{1}{T^2}\sum_{j_1,\ldots,j_4=0}^{\infty}\sum_{r,t_1,t_2\in\mathcal{A}_1}\lambda_M(r)^2 M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1+r-j_2)M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4).$$

The other 23 cases are defined similarly (different orderings of $t_1-j_1,\ldots,t_2+r-j_4$), such that we have $I_{\mathcal{A}}=\sum_{j=1}^{24}I_{\mathcal{A}_j}$. We start by bounding $I_{\mathcal{A},1}$. Since $t_1-j_1>t_1+r-j_2>t_2-j_3>t+r-j_4$, it is easy to see that $M_{j_1}^{(x)}(t_1-j_1)\sum_{r<j_2-j_1}\lambda_M(r)^2M_{j_2}^{(y)}(t_1+r-j_2)\sum_{t_2<t_1-j_1+j_3}M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4)$ is a martingale over t_1, $M_{j_2}^{(y)}(t_1+r-j_2)\sum_{t_2<t_1-j_1+j_3}M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4)$ is a martingale over r and $\{M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4)\}$ is a martingale over t_2. Thus by using Burkhölder's inequality together with Hölder's inequality three times, for any $q\geq 2$ we have

$$\begin{aligned}\|I_{\mathcal{A},1}\|_q &= \frac{1}{T^2}\sum_{j_1,\ldots,j_4=0}^{\infty}\Big(\sum_{r,t_1,t_2}\lambda_M(r)^2\|M_{j_1}^{(x)}(t_1-j_1)\|_{4q}^2\|M_{j_2}^{(y)}(t_1+r-j_2)\|_{4q}^2 \\ &\quad \|M_{j_3}^{(x)}(t_2-j_3)\|_{4q}^2\|M_{j_4}^{(y)}(t_2+r-j_4)\|_{4q}^2\Big)^{1/2}.\end{aligned}$$

Thus by using (16) we have that $\|I_{\mathcal{A},1}\|_q=O(\frac{M^{1/2}}{T})$ and by the same argument we have $I_{\mathcal{A},j}=O(\frac{M^{1/2}}{T})$ (for $2\leq j\leq 24$). Therefore, altogether this gives $\|I_{\mathcal{A}}\|_q=O(\frac{M^{1/2}}{T})$. We now bound $I_{\mathcal{B}}$. We first define the random variable

$$\begin{aligned}&A_{j_1,j_2;i}^{(x,y)}(t_1-j_1-i) = \\ &\mathbb{E}(M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1-j_1)|\mathcal{F}_{t_1-j_1-i})-\mathbb{E}(M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1-j_1)|\mathcal{F}_{t_1-j_1-i}).\end{aligned}$$

To bound $\|A^{(x,y)}_{j_1,j_2;i}(t_1-j_1-i)\|_q$, we repeatedly use Ibragimov's inequality and (16) to give

$$\begin{aligned}\|A^{(x,y)}_{j_1,j_2;i}(t_1-j_1-i)\|_q &\leq 2\|\mathbb{E}(\overline{M^{(x)}_{j_1}(t_1-j_1)M^{(y)}_{j_2}(t_1-j_1)}|\mathcal{F}_{t_1-j_1-i})\| \\ &\leq C\alpha(i)\|M^{(x)}_{j_1}(t_1-j_1)M^{(y)}_{j_2}(t_1-j_1)\|_q \leq C\alpha(i)\alpha(j_1)\alpha(j_2). \quad (17)\end{aligned}$$

This gives the representation

$$\overline{M^{(x)}_{j_1}(t_1-j_1)M^{(y)}_{j_2}(t_1-j_1)} = \sum_i A^{(x,y)}_{j_1,j_2;i}(t_1-j_1-i).$$

Substituting the above representation into $I_{\mathcal{B}}$ gives

$$\begin{aligned} I_{\mathcal{B}} &= \frac{1}{T^2}\sum_{j_1,\ldots,j_4,i_1,i_2=0}^{\infty}\sum_{t_1,t_2\in\mathcal{B}}\lambda_M(j_1-j_2)^2\big[A^{(x,y)}_{j_1,j_2;i_1}(t_1-j_1-i)A^{(x,y)}_{j_3,j_4;i_2}(t_2-j_3-i_2) \\ &\quad -\mathbb{E}(A^{(x,y)}_{j_1,j_2;i_1}(t_1-j_1-i)A^{(x,y)}_{j_3,j_4;i_2}(t_2-j_3-i_2))\big] \\ &:= I_{\mathcal{B},1}+I_{\mathcal{B},2}+I_{\mathcal{B},3},\end{aligned}$$

where

$$I_{\mathcal{B},1} := \frac{1}{T^2}\sum_{j_1,\ldots,j_4,i_1,i_2=0}^{\infty}\sum_{t_1-j_1-i_1>t_2-j_3-i_2}\lambda_M(j_1-j_2)^2A^{(x,y)}_{j_1,j_2;i_1}(t_1-j_1-i)A^{(x,y)}_{j_3,j_4;i_2}(t_2-j_3-i_2)$$

$$I_{\mathcal{B},2} := \frac{1}{T^2}\sum_{j_1,\ldots,j_4,i_1,i_2=0}^{\infty}\sum_{t_1-j_1-i_1<t_2-j_3-i_2}\lambda_M(j_1-j_2)^2A^{(x,y)}_{j_1,j_2;i_1}(t_1-j_1-i)A^{(x,y)}_{j_3,j_4;i_2}(t_2-j_3-i_2)$$

$$I_{\mathcal{B},3} := \frac{1}{T^2}\sum_{j_1,\ldots,j_4,i_1,i_2=0}^{\infty}\sum_{t_1-j_1-i_1=t_2-j_3-i_2}\lambda_M(j_1-j_2)^2\overline{A^{(x,y)}_{j_1,j_2;i_1}(t_1-j_1-i)A^{(x,y)}_{j_3,j_4;i_2}(t_2-j_3-i_2)}.$$

Using similar techniques to those used to bound $\|I_{\mathcal{A},1}\|_q$, Burkhölder's and Hölder's inequalities twice on $\|I_{\mathcal{B},1}\|_q$, together with (17), we obtain the bound $\|I_{\mathcal{B},1}\|_q = O(\frac{1}{T})$. A similar argument can be used for $\|I_{\mathcal{B},2}\|_q = O(\frac{1}{T})$. To bound $\|I_{\mathcal{B},3}\|_q$, we need to decompose

$$A^{(x,y)}_{j_1,j_2;i_1}(t_1-j_1-i)A^{(x,y)}_{j_3,j_4;i_2}(t_2-j_3-i_2)-\mathbb{E}(A^{(x,y)}_{j_1,j_2;i_1}(t_1-j_1-i)A^{(x,y)}_{j_3,j_4;i_2}(t_2-j_3-i_2)),$$

into the sum of martingale differences, using this martingale decomposition we can use the same argument as those used above to obtain $\|I_{\mathcal{B},3}\| = O(\frac{1}{T^{3/2}})$. Therefore, altogether we have $\|I_{\mathcal{B}}\|_q = O(\frac{1}{T})$. Now by using similiar arguments and repeated decompositions into martingale differences we can show that $\|I_{\mathcal{C}}\|_q,\ldots,\|I_{\mathcal{F}}\|_q = O(\frac{M^{1/2}}{T^{3/2}})$. Thus we have shown that $I_{\mathcal{A}}$ is the dominating term in $\mathcal{Q}_T-\mathbb{E}(\mathcal{Q}_T)$. Since $\mathcal{S}_T = \int I_{\mathcal{A}}dF_0(x)dF_0(y)$ we have obtained the desired result. $\square$

To prove the result we use the martingale central limit theorem on

$$\mathcal{S}_T = \frac{1}{T^2}\int \sum_{j_1,\ldots,j_4=0}^{\infty} \sum_{t_1,r,t_2\in\mathcal{A}} \lambda_M(r)^2 M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1+r-j_2)M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4)dF_0(x)dF_0(y).$$

To do this, we use the same decompositions of $I_{\mathcal{A}}$, as that used in the proof of Theorem A.1. We set $\mathcal{S}_{T,i} := I_{\mathcal{A},i}$, recalling that

$$\mathcal{S}_{T,i} = \frac{1}{T^2}\sum_{j_1,\ldots,j_4=0}^{\infty} \sum_{r,t_1,t_2\in\mathcal{A},i} \lambda_M(r)^2 M_{j_1}^{(x)}(t_1-j_1)M_{j_2}^{(y)}(t_1+r-j_2)M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4),$$

where $\mathcal{A}_i$ is an ordering of $\{t_1-j_1, t_1+r-j_2, t_2-j_3, t_2+r-j_4\}$. We show that $\mathcal{S}_{T,i}$ can be written as the sum of martingale differences. First consider $\mathcal{S}_{T,1}$, this can be written as $\mathcal{S}_{T,1} = \frac{1}{T^2}\sum_{k=1}^{T} U_{k,1}$, where with a change of variables we have

$$U_{k,1} = \int \sum_{j_1=0}^{T-k} M_{j_1}(k) \sum_{j_2,j_3,j_4} \sum_{r,t_1\in\tilde{\mathcal{A}}_{k,1}} \lambda_M(r)^2 M_{j_2}^{(y)}(k+j_1+r-j_2)M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4)dF_0(x)dF_0(y)$$

and $\tilde{\mathcal{A}}_{k,1} = \{(r,t_2) \text{ such that } (k > k+j_1+r-j_2 > t_2-j_3 > t_2+r-j_4)\}$. Using a similar argument we can decompose $\mathcal{S}_{T,i}$ as $\mathcal{S}_{T,i} = \frac{1}{T^2}\sum_{k=1}^{T} U_{k,i}$ (and $U_{k,i}$ is defined similar to above). Therefore, altogether $\mathcal{S}_T$ is the sum of martingale differences, where $\mathcal{S}_T = \frac{1}{T^2}\sum_{k=1}^{T}\sum_{i=1}^{24} U_{k,i}$, and $\sum_{i=1}^{24} U_{k,i} \in \sigma(X_k, X_{k-1},\ldots)$ are the martingale differences. Therefore under the conditions in Theorem A.1 we have

$$\mathcal{Q}_T - \mathbb{E}(\mathcal{Q}_T) = \mathcal{S}_T + O_p(\frac{1}{T}+\frac{M^{1/2}}{T^2}) = \frac{1}{T^2}\sum_{k=1}^{T}\sum_{i=1}^{24} U_{k,i} + O_p(\frac{1}{T}+\frac{M^{1/2}}{T^2}).$$

These approximations will allow us to use the martingale central limit theorem to prove asymptotic normality, which requires the following lemma.

Lemma A.3 *Suppose that Assumption 3.1 holds. Then for all* $1 \leq i \leq 24$ *and* $1 \leq k \leq T$ *we have*

$$\|\sum_{i=1}^{24} U_{k,i}\|_q = O(T^{1/2}M^{1/2}) \tag{18}$$

$$\frac{1}{T^2M}\sum_{k=1}^{T}\mathbb{E}(\sum_{i=1}^{24} U_{k,i}^2) \to \frac{4}{M}\int\int \Delta_M(\theta_1-\theta_2)^2 \prod_{i=1}^{2} G(x_1,x_2;\theta_i)G(y_1,y_2;\theta_i)d\theta_i \prod_{j=1}^{2} dF_0(x_j)dF_0(y_j) \tag{19}$$

and

$$\frac{1}{T^2 M}\sum_{k=1}^{T}\Big[\mathbb{E}\Big((\sum_{i=1}^{24}U_{k,i})^2|\mathcal{F}_{k-1}\Big)-\mathbb{E}\big(\sum_{i=1}^{24}U_{k,i}\big)^2\Big]\overset{\mathcal{P}}{\to}0. \tag{20}$$

PROOF. To prove the result we concentrate on $U_{k,1}$, a similar proof applies to the other terms. By using the Hölder inequality, for any $q\geq 2$, we obtain

$$\begin{aligned}\|U_{k,1}\|_q &\leq \int\sum_{j_1=0}^{T-k}\|M_{j_1}(k)\|_{4q}\Big\|\sum_{j_2,j_3,j_4}\sum_{r,t_1\in\tilde{\mathcal{A}}_{k,1}}\lambda_M(r)^2 M_{j_2}^{(y)}(k+j_1+r-j_2)\times\\ &\qquad M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4)\Big\|_{4q/3}dF_0(x)dF_0(y).\end{aligned}$$

Now by repeated use of Burkhölder's inequality we have $\|U_{k,1}\|_q=O(M^{1/2}T^{1/2})$, using a similar method we obtain a similar bound for $\|U_{k,i}\|_q$, this gives (18).

The proof of (19) follows from the proof of Theorem A.1 (noting that the asymptotic variance of $\mathcal{Q}_{\mathcal{T}}$ is determined by the variance of $\mathcal{S}_T$).

To prove (20), we consider only the $U_{k,1}$ (the proof involving the other terms in similar). For brevity we write $U_{k,1}$ as

$$U_{k,1}=\int\sum_{j_1=0}^{T-k}M_{j_1}^{(x)}(k)N_{j_1,k-1,1}^{(x,y)}dF_0(x)dF_0(y),$$

where

$$N_{j_1,k-1,1}^{(x,y)} = \sum_{j_2,j_3,j_4}\sum_{r,t_1\in\tilde{\mathcal{A}}_{k,1}}\lambda_M(r)^2 M_{j_2}^{(y)}(k+j_1+r-j_2)M_{j_3}^{(x)}(t_2-j_3)M_{j_4}^{(y)}(t_2+r-j_4).$$

Noting that $N_{j_1,k-1,1}^{(x,y)}\in\mathcal{F}_{k-1}$ we have

$$\begin{aligned}&\frac{1}{T^2M}\sum_{k=1}^{T}\big(\mathbb{E}\big(U_{k,1})^2|\mathcal{F}_{k-1}\big)-\mathbb{E}\big(U_{k,1}\big)^2\big) =\\ &\frac{1}{T^2M}\sum_{k=1}^{T}\int\sum_{j_1,j_2=0}^{T-k}\big(\mathbb{E}(M_{j_1}^{(x)}(k)M_{j_2}^{(x_1)}(k)|\mathcal{F}_{k-1})-\mathbb{E}(M_{j_1}^{(x)}(k)M_{j_2}^{(x_2)}(k))\big)N_{j_1,k-1,1}^{(x_1,y_2)}N_{j_1,k-1,1}^{(x_2,y_2)}\prod_{i=1}^{2}dF_0(x_i)dF_0(y_i)\\ &+\frac{1}{T^2M}\sum_{k=1}^{T}\int\sum_{j_1,j_2=0}^{T-k}\mathbb{E}(M_{j_1}^{(x)}(k)M_{j_2}^{(x_2)}(k))\big)\big(N_{j_1,k-1,1}^{(x_1,y_2)}N_{j_1,k-1,1}^{(x_2,y_2)}-\mathbb{E}(N_{j_1,k-1,1}^{(x_1,y_2)}N_{j_1,k-1,1}^{(x_2,y_2)})\big)\prod_{i=1}^{2}dF_0(x_i)dF_0(y_i).\end{aligned}$$

Now by using similar methods to the iterative martingale methods detailed in the proof of Theorem A.1, we can show that the $\|\cdot\|_q$-norm ($q\geq 2$) of the above converges to zero, thus we

have (20).

PROOF of Theorem 3.2 Using the above we have $\mathcal{Q}_T = \frac{1}{T^2}\sum_{k=1}^{T}\sum_{i=1}^{24} U_{k,i} + O_p(\frac{1}{T} + \frac{M^{1/2}}{T^2})$, thus $\mathcal{Q}_T - \mathbb{E}(\mathcal{Q}_T)$ can be written as the sum of martingales plus a smaller order term. Therefore to prove asymptotic normality of $\mathcal{Q}_T$ we can use the martingale central limit, for this we need to verify (a) the conditional Lindeberg condition $\frac{1}{MT^2}\sum_{k=1}^{T}\mathbb{E}(|\sum_{i=1}^{24} U_{k,i}|^2 I(\frac{1}{M^{1/2}T}|\sum_{i=1}^{24} U_{k,i}| > \varepsilon)|\mathcal{F}_{k-1}) \stackrel{\mathcal{P}}{\to} 0$ for all $\varepsilon > 0$, (b) that $\frac{1}{MT^2}\sum_{k=1}^{T}\mathbb{E}(|\sum_{i=1}^{24} U_{k,i}|^2|\mathcal{F}_{k-1}) - \frac{T^2}{M}\mathrm{var}(Q_T) \stackrel{\mathcal{P}}{\to} 0$.

To verify the conditional Lindeberg condition, we observe that the Cauchy-Schwartz and Markov's inequalities give

$$\frac{1}{MT^2}\sum_{k=1}^{T}\mathbb{E}\big(|\sum_{i=1}^{24} U_{k,i}|^2 I(\frac{1}{M^{1/2}T}|\sum_{i=1}^{24} U_{k,i}| > \varepsilon)|\mathcal{F}_{k-1}\big) \leq \frac{1}{\varepsilon M^2T^4}\sum_{k=1}^{T}\mathbb{E}(|\sum_{i=1}^{24} U_{k,i}|^4|\mathcal{F}_{k-1}) := B_T.$$

By using (18) the expectation of the above is $\mathbb{E}(B_T) = O(\frac{1}{T})$. As B_T is a non-negative random variable, this implies $B_T \stackrel{\mathcal{P}}{\to} 0$ as $T \to \infty$. Thus we have shown that the Lindeberg condition is satisfied. To prove (b) we note that

$$\begin{aligned}
&\frac{1}{MT^2}\sum_{k=1}^{T}\mathbb{E}\Big(|\sum_{i=1}^{24} U_{k,i}|^2|\mathcal{F}_{k-1}\Big) - \frac{T^2}{M}\mathrm{var}(Q_T) \\
= &\frac{1}{MT^2}\sum_{k=1}^{T}\Big[\mathbb{E}\big(|\sum_{i=1}^{24} U_{k,i}|^2|\mathcal{F}_{k-1}\big) - \mathbb{E}(|\sum_{i=1}^{24} U_{k,i}|^2)\Big] + \frac{1}{MT^2}\sum_{k=1}^{T}\mathbb{E}(|\sum_{i=1}^{24} U_{k,i}|^2) - \frac{T^2}{M}\mathrm{var}(Q_T).
\end{aligned}$$

By using (19) and (20) the above converges to zero in probability. Thus we have verified the conditions of the martingale central limit theorem and we have the desired result. □

A.2 Proof of Theorem 3.3

As the limiting distribution of $\mathcal{Q}_T$ is determined by $\mathcal{Q}_{T,2}$, we rewrite $\mathcal{Q}_{T,2}$ in such a way that the same methods used to prove Theorem 2 in Lee and Subba Rao (2010), can be used to obtain the limiting distribution. We observe that

$$\begin{aligned}
\mathcal{Q}_{T,2} &= \frac{2}{T}\Re\int\sum_{k}\Lambda_T(x,y;\omega_k)\big\{J_T(x;\omega_k)\overline{J}_T(y;\omega_k) - \mathbb{E}(J_T(x;\omega_k)\overline{J}_T(y;\omega_k))\big\}dF_0(x)dF_0(y) \\
&= \int\frac{2}{T}\sum_{t,\tau}\lambda_M(t-\tau)^2 D_{t-\tau,T}(x,y)(Z_t(x)Z_\tau(y) - \mathbb{E}(Z_t(x)Z_\tau(y)))dF_0(x)dF_0(y) \\
&= \int\frac{2}{T}\sum_{t,\tau}\lambda_M(t-\tau)^2 D_{t-\tau,T}(x,y)(\tilde{Z}_T(x)\tilde{Z}_\tau(y) - C_r(x,y))dF_0(x)dF_0(y) + O_p(\frac{1}{T}),
\end{aligned}$$

where $\Lambda_T(x,y;\omega_s)=\sum_r \lambda_M(r)^2(\frac{T-|r|}{T})[C_{r,1}(x,y)-C_{r,0}(x,y)]\exp(ir\omega_k)$, $D_{r,T}(x,y)=(\frac{T-|r|}{T})[C_{r,1}(x,y)-C_{r,0}(x,y)]$ and $\tilde{Z}_t(x)=I(X_t\leq x)-F_1(x)$.

PROOF of Theorem 3.3 Now we observe that under the stated assumptions of the theorem we have that the quantile covariances under the null decay at the rate $\sup_{x,y}|C_{r,0}(x,y)|\leq K|r|^{-(2+\delta)}$ (for some $\delta>0$) and $\sup_{x,y}|C_{r,1}(x,y)|\leq K|r|^{-s}$ (for some $s>2$). Thus by definition of $D_{r,T}(\cdot)$, we have $\sup_{x,y}|\lambda_M(r)D_{r,T}(x,y)|\leq K|r|^{-\min(2+\delta,s)}$. Thus we can write $\mathcal{Q}_{T,2}$ as

$$\mathcal{Q}_{T,2} = \int \frac{2}{T}\sum_{t,\tau}\lambda_M(t-\tau)^2 D_{t-\tau,T}(x,y)(\tilde{Z}_T(x)\tilde{Z}_\tau(y)-\mathbb{E}(\tilde{Z}_T(x)\tilde{Z}_\tau(y)))dF_0(x)dF_0(y)+O_p(\frac{1}{T}),$$

where we observe that terms where $|t-\tau|>2M$, are zero. Thus using the Bernstein blocking arguments for quadratic forms used to prove Theorem 2, Lee and Subba Rao (2011), we can show asymptotic normality of the above. This proves (6). Finally to prove (7), we note that $\mathcal{Q}_T=\mathcal{Q}_{T,2}+E_{T,2}+O_p(\frac{M^{1/2}}{T}+\frac{M}{T}+\frac{1}{M^{s-1}})$, by using (6), this immediately leads to (7). □

References

Anderson, T. W. (1993). Goodness of fit tests for spectral distributions. *Ann. Statist.*, *21*(2), 830–847.

Berkowitz, J., & Diebold, F. X. (1998). Bootstrapping multivariate spectra. *Review of Economics and Statistics*, *80*(4), 664–666.

Brillinger, D. (1981). *Time series, data analysis and theory.* San Francisco: SIAM.

Chen, W. W., & Deo, R. S. (2004). A generalized Portmanteau goodness-of-fit test for time series models. *Econometric Theory*, *20*(2), 382–416.

Dette, H., Hallin, M., Kley, T., & Volgushav, S. (2011). Of copulas, quantile, ranks and spectra: and l_1 approach to spectral analysis. *arxiv:1111.7205.*

Dette, H., & Paparoditis, E. (2009). Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(4), 831–857.

Franke, J., & Härdle, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.*, *20*(1), 121–145.

Hagemann, A. (2011). Robust spectral analysis. *arxiv:111.1965v1, Arxiv, e-prints.*

Hallin, M., & Puri, M. (1992). Rank tests for time series: a survey. In P. I. New directions in time series analysis (Ed.), (Vol. 45). New York: Springer.

Hoeffding, W. (1948). A non-parametric test of independence. *Ann. Math. Statistics*, *19*, 546–557.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica*, *64*(4), 837–864.

Hong, Y. (1998). Testing for pairwise serial independence via the empirical distribution function. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *60*(2), 429–453.

Hong, Y. (2000). Generalized spectral tests for serial dependence. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *62*(3), 557–574.

Hong, Y., & Lee, T.-H. (2003). Diagnostic checking for the adequacy of nonlinear time series models. *Econometric Theory*, *19*(6), 1065–1121.

Hurvich, C., & Zeger, S. (1987). Frequency domain bootstrap methods for time series. *New York University, Graduate School of Business Adminstration*.

Lee, J., & Subba Rao, S. (2010). A note on general quadratic forms of nonstationary time series. *Submitted*.

Li, T.-H. (2008). Laplace periodograms for time series analysis. *Journal of the American Statistical association*, *103*, 757-768.

Neumann, M. H., & Paparoditis, E. (2008). Goodness-of-fit tests for Markovian time series models: central limit theory and bootstrap approximations. *Bernoulli*, *14*(1), 14–46.

Priestley, M. B. (1981). *Spectral analysis and time series.* London: Academic Press.

Skaug, H. J., & Tjøstheim, D. (1993). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika*, *80*(3), 591–602.

Tsay, R. S. (2005). *Analysis of financial time series* (2nd ed.). Wiley-Interscience.

Zivot, E. (2009). Practical issues in the analysis of univariate garch models. In T. G. Andersen, R. Davis, J.-P. Kreiss, & T. Mikosch (Eds.), *Handbook of financial time series.* springer.

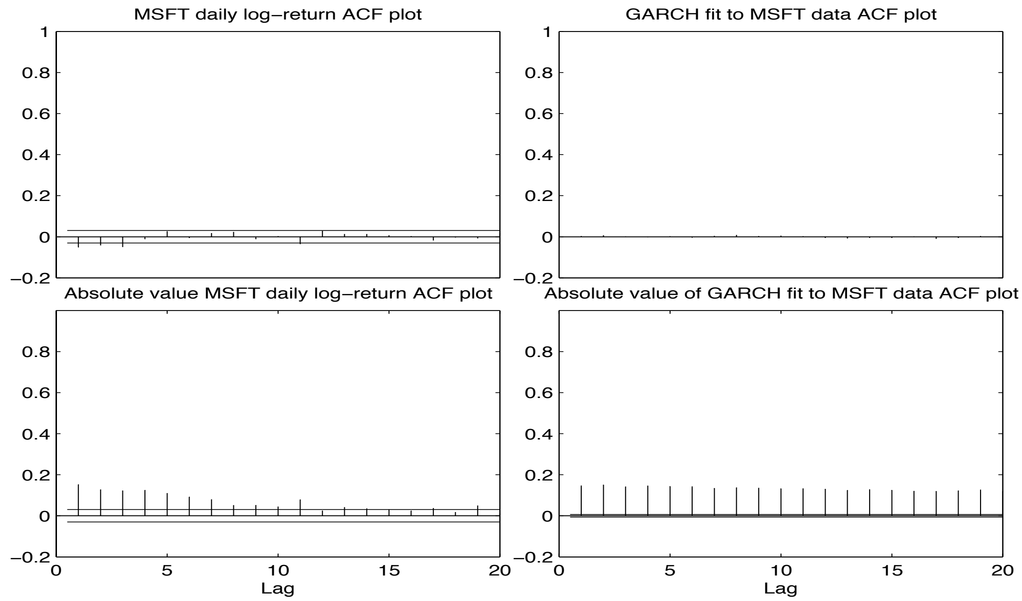

Figure 1: The ACF plots of $\{X_t\}$ and $\{|X_t|\}$ of the MSFT and the corresponding GARCH model

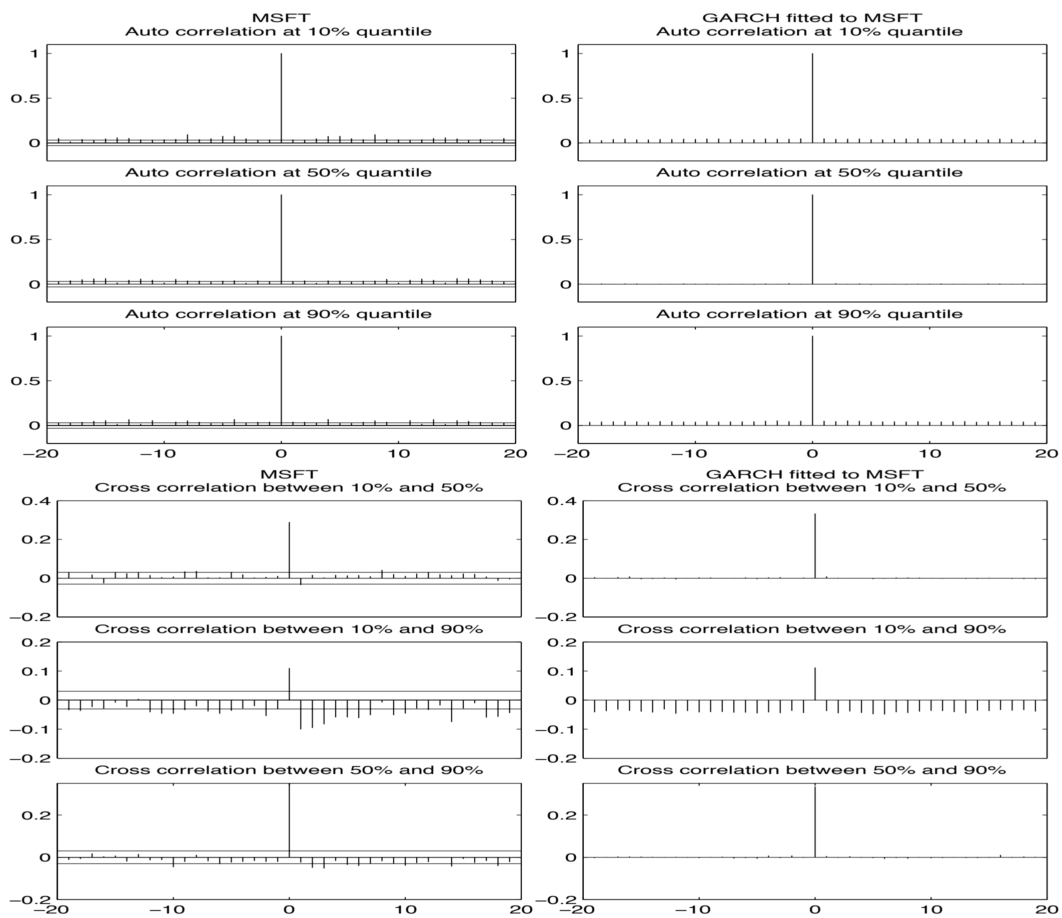

Figure 2: The quantile covariance of the MSFT and the corresponding GARCH

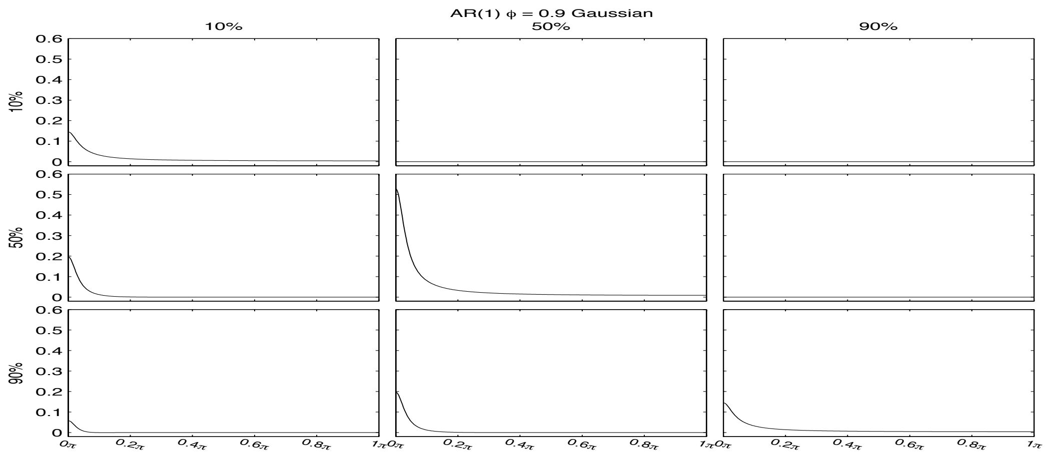

Figure 3: The quantile spectral density of $X_t = 0.9X_{t-1} + Z_t$

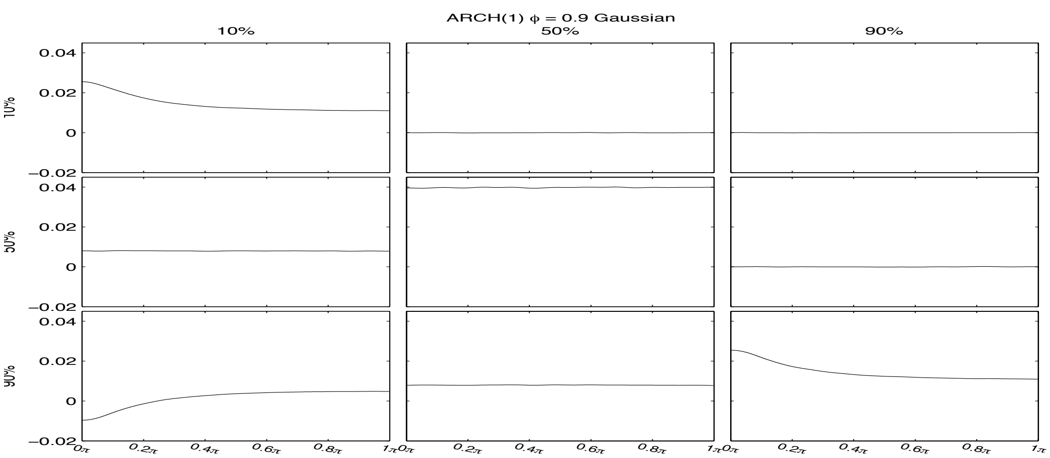

Figure 4: The quantile spectral density of $X_t = \sigma_t Z_t$, where $\sigma_t^2 = 1/1.9 + 0.9X_{t-1}^2$

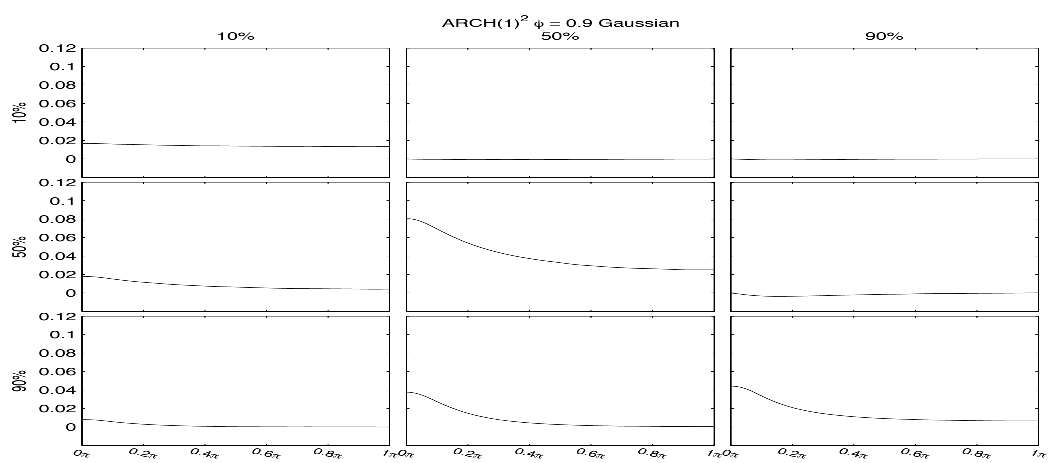

Figure 5: The quantile spectral density of $X_t^2 = \sigma_t^2 Z_t^2$, where $\sigma_t^2 = 1/1.9 + 0.9X_{t-1}^2$

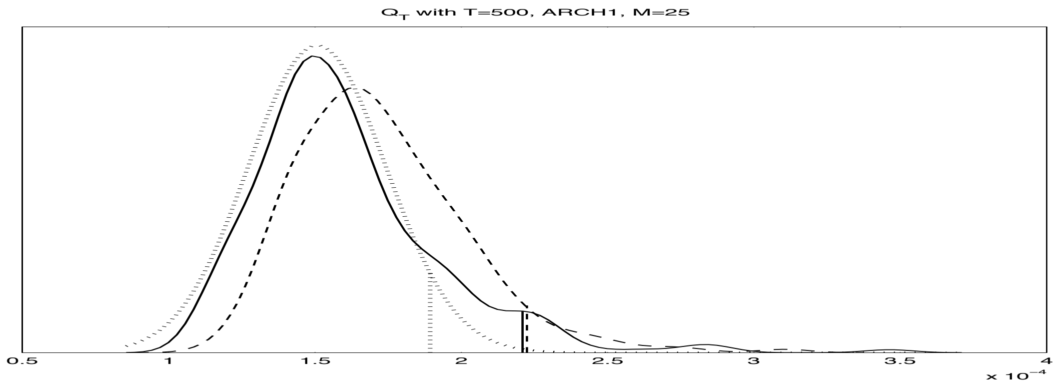

Figure 6: The fine line is the standard normal (with the 5% rejection line), the thick solid line is the finite sample density of the test statistic (with 5% rejection region) and the thick dashed line is the bootstrap approximation (with 5% rejection region).

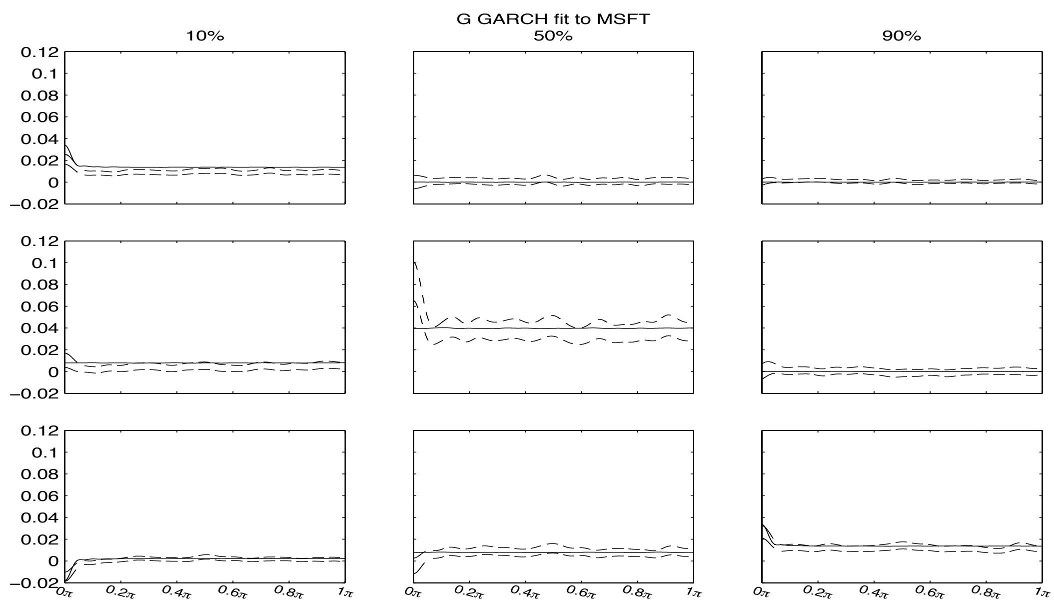

Figure 7: The quantile spectral density of the fitted GARCH(1, 1) model using Microsoft data with the confidence intervals

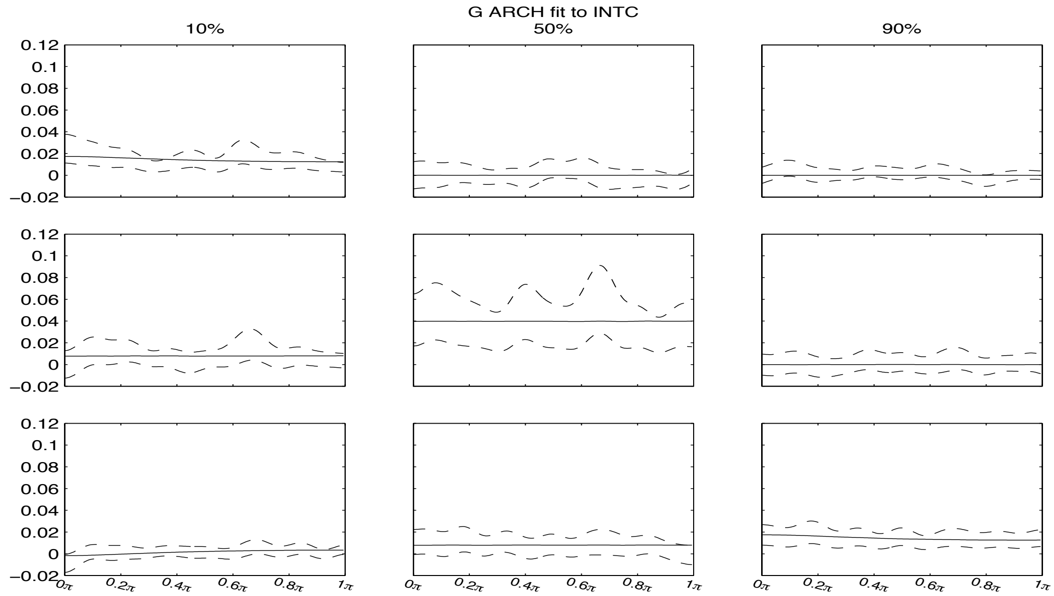

Figure 8: The quantile spectral density of the fitted ARCH(1) from Intel data with the confidence intervals

Table 2: $H_0 : AR(1)$, $H_A : ARCH$ $T = 100$

$T = 100$		$\alpha = 0.1$				$\alpha = 0.05$			
		Bootstrap		Normal		Bootstrap		Normal	
a	M	H_0	H_A	H_0	H_A	H_0	H_A	H_0	H_A
0.3	11	0.052	1	0.076	1	0.021	0.972	0.054	1
	16	0.04	0.869	0.062	0.971	0.011	0.262	0.04	0.854
	21	0.048	0.386	0.064	0.561	0.021	0.106	0.043	0.348
	25	0.021	0.071	0.048	0.229	0.014	0.016	0.029	0.12
0.4	11	0.048	1	0.082	1	0.02	1	0.055	1
	16	0.043	1	0.059	1	0.013	0.939	0.041	1
	21	0.046	0.932	0.066	0.997	0.011	0.416	0.046	0.929
	25	0.036	0.582	0.055	0.832	0.01	0.124	0.037	0.598
0.5	11	0.046	1	0.073	1	0.015	1	0.052	1
	16	0.049	1	0.078	1	0.027	1	0.045	1
	21	0.046	1	0.06	1	0.015	0.985	0.037	1
	25	0.047	1	0.062	1	0.015	0.397	0.043	1
0.55	11	0.041	1	0.096	1	0.018	1	0.057	1
	16	0.045	1	0.066	1	0.017	1	0.046	1
	21	0.065	1	0.06	1	0.034	1	0.034	1
	25	0.045	1	0.051	1	0.024	1	0.032	1

Table 3: $H_0 : AR(1),\ H_A : ARCH(1)\ T = 500$

$T = 500$		$\alpha = 0.1$				$\alpha = 0.05$			
		Bootstrap		Normal		Bootstrap		Normal	
a	M	H_0	H_A	H_0	H_A	H_0	H_A	H_0	H_A
0.3	14	0.053	1	0.098	1	0.024	1	0.063	1
	21	0.064	1	0.082	1	0.023	1	0.052	1
	28	0.06	1	0.093	1	0.024	1	0.062	1
	35	0.07	1	0.086	1	0.033	1	0.062	1
0.4	14	0.043	1	0.092	1	0.014	1	0.064	1
	21	0.058	1	0.092	1	0.015	1	0.056	1
	28	0.066	1	0.094	1	0.03	1	0.061	1
	35	0.073	1	0.087	1	0.032	1	0.052	1
0.5	14	0.031	1	0.105	1	0.018	1	0.072	1
	21	0.059	1	0.079	1	0.03	1	0.05	1
	28	0.076	1	0.111	1	0.046	1	0.069	1
	35	0.053	1	0.086	1	0.022	1	0.055	1
0.55	14	0.038	1	0.107	1	0.014	1	0.077	1
	21	0.056	1	0.108	1	0.021	1	0.067	1
	28	0.071	1	0.103	1	0.032	1	0.06	1
	35	0.051	1	0.089	1	0.026	1	0.06	1

Table 4: $H_0 : ARCH(1),\ H_A = AR(1)\ T = 100$

$T = 100$		$\alpha = 0.1$				$\alpha = 0.05$			
		Bootstrap		Normal		Bootstrap		Normal	
a	M	H_0	H_A	H_0	H_A	H_0	H_A	H_0	H_A
0.3	11	0.039	0.994	0.08	0.997	0.022	0.984	0.051	0.995
	16	0.043	0.978	0.086	0.991	0.009	0.925	0.055	0.983
	21	0.045	0.98	0.07	0.99	0.016	0.934	0.051	0.983
	25	0.026	0.939	0.059	0.976	0.011	0.895	0.045	0.965
0.4	11	0.046	1	0.086	1	0.012	0.999	0.053	1
	16	0.049	0.993	0.092	0.999	0.014	0.988	0.062	0.996
	21	0.03	0.994	0.07	0.997	0.017	0.983	0.046	0.997
	25	0.038	0.994	0.083	0.997	0.024	0.982	0.059	0.994
0.5	11	0.054	1	0.107	1	0.024	1	0.067	1
	16	0.063	1	0.098	1	0.03	1	0.066	1
	21	0.051	1	0.083	1	0.022	1	0.061	1
	25	0.028	0.997	0.06	0.998	0.012	0.995	0.043	0.998
0.55	11	0.074	1	0.113	1	0.03	1	0.081	1
	16	0.056	1	0.087	1	0.02	1	0.054	1
	21	0.065	1	0.08	1	0.038	1	0.057	1
	25	0.067	1	0.088	1	0.03	1	0.065	1

Table 5: $H_0 : ARCH(1)$ and $H_A : AR(1)$ $T = 500$

$T = 500$		$\alpha = 0.1$				$\alpha = 0.05$			
		Bootstrap		Normal		Bootstrap		Normal	
a	M	H_0	H_A	H_0	H_A	H_0	H_A	H_0	H_A
0.3	14	0.072	1	0.09	1	0.025	1	0.059	1
	21	0.062	1	0.094	1	0.032	1	0.059	1
	28	0.067	1	0.097	1	0.024	1	0.062	1
	35	0.076	1	0.101	1	0.026	1	0.073	1
0.4	14	0.045	1	0.097	1	0.022	1	0.059	1
	21	0.075	1	0.105	1	0.03	1	0.077	1
	28	0.06	1	0.111	1	0.024	1	0.07	1
	35	0.085	1	0.12	1	0.041	1	0.086	1
0.5	14	0.053	1	0.129	1	0.032	1	0.079	1
	21	0.1	1	0.121	1	0.054	1	0.082	1
	28	0.111	1	0.124	1	0.071	1	0.085	1
	35	0.066	1	0.117	1	0.029	1	0.075	1
0.55	14	0.099	1	0.143	1	0.047	1	0.104	1
	21	0.074	1	0.119	1	0.042	1	0.083	1
	28	0.078	1	0.11	1	0.037	1	0.072	1
	35	0.082	1	0.119	1	0.037	1	0.085	1